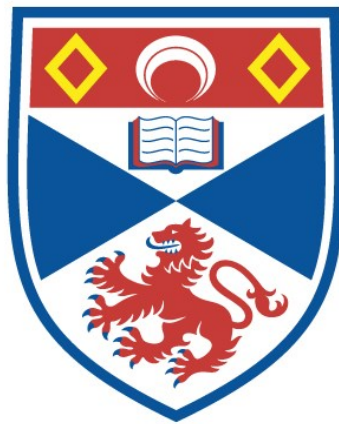


PERSONAL IDENTITY AND PRACTICAL REASON

Patrik Alexander Hummel

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2018

Full metadata for this item is available in
St Andrews Research Repository
at:
<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:
<http://hdl.handle.net/10023/16607>

This item is protected by original copyright

Personal Identity and Practical Reason

Patrik Alexander Hummel



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of PhD
at the
University of St Andrews

18th October 2017

ST ANDREWS AND STIRLING
GRADUATE PROGRAMME FOR
PHILOSOPHY



University of
St Andrews

UNIVERSITY of
STIRLING



PHD DISSERTATION

**Personal Identity and
Practical Reason**

Patrik Alexander Hummel

18 October 2017

1. Candidate's declarations:

I, Patrik Alexander Hummel, hereby certify that this thesis, which is approximately 74000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in May, 2013 and as a candidate for the degree of Doctor of Philosophy in May 2013; the higher study for which this is a record was carried out in the University of St Andrews between 2013 and 2017.

Date 18 October 2017 signature of candidate

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date signature of supervisor

3. Permission for publication: (to be signed by both candidate and supervisor)

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

PRINTED COPY

a) No embargo on print copy

ELECTRONIC COPY

a) No embargo on electronic copy

Date 18 October 2017 signature of candidate

signature of supervisor

Abstract

In this thesis, I argue that the interdependence between personal identity and practical concerns is overstated. In paradigmatic places where philosophers and common sense suggest that personal identity constrains how we should reason and care, or vice versa, the two spheres are in fact neutral to each other. I defend this claim by considering four specific cases. First, a rough characterization of the distinction between the complex and the simple view is that the former takes personal identity to consist in other relations, whereas the latter does not. I argue that the *extreme claim* according to which the complex view fails to give reasons for future-directed concern can be resisted. We maintain forward-looking attitudes and projects not because someone will be *us*, but because we relate to future selves in other, more important ways. Second, I argue that intuitions in a range of popular imaginary cases are contaminated by practical concerns whose relevance for personal identity is far from straightforward. Third, I argue that on a closer look, the complex versus simple distinction is confused. It thus cannot be what grounds differences in judgements on what matters. Debates about personal identity should be framed in terms of better understood notions. Finally, I argue that it is not a constraint on rational transformative choice that decision-maker and transforming individual are identical. Moreover, whether we are deciding for ourselves or for others—the importance of informed consent for transformative treatments is not diminished by the decision-maker’s failure to projectively imagine the outcomes.

Acknowledgements

First and foremost, I thank my parents Nairy and Günter, and my siblings Sebastian, Felix, Timon, and Chiara as well as Gudrun Gramlich and Annette Hummel-Grether. A special thanks goes to Lorianna Clarke-Alexander, who was always there for me throughout this process.

I was extremely lucky to have Katherine Hawley as my supervisor. I am grateful for her continuous support and encouragement with regards to philosophical thinking and writing as well as its impact outside academia. Steinvör Thöll Árnadóttir and Patrick Greenough and my annual reviewers Jessica Brown and Simon Prosser provided thorough feedback on several chapters. I thank Sònia Roca Royes for her generous help with numerous applications. I will miss the unique atmosphere at Arché! My colleagues were a constant source of support and inspiration.

I am grateful for funding from the German Academic Exchange Service (DAAD) and the Royal Institute of Philosophy as well as a fee waiver from the University of St Andrews. I also thank the University of St Andrews and my supervisors for allowing me to take a leave of absence for an internship with the Global Health Ethics Unit of the World Health Organization (WHO) in Geneva. The internship was supported by a Carlo Schmid Fellowship from DAAD. At WHO, I received invaluable supervision from Abha Saxena, Andreas Reis, Maria Magdalena Guraiib, and Vânia de la Fuente-Núñez, which gave my research several new directions.

A few months later, I spent one semester at the School of Philosophy at Australian National University (ANU). Brian Garrett provided helpful feedback on a number of ideas defended in this dissertation. The whole ANU Philosophy community made my stay memorable.

Contents – Compact

Introduction	1
1 The Complex Versus Simple Distinction	9
1.1 Positive Reasons	11
1.2 Negative Reasons	17
1.3 Conclusion	24
2 The Extreme Claim	27
2.1 Practical Objections	29
2.2 The Extreme And The Moderate Claim	33
2.3 A Reductio Against The Complex View	38
2.4 Response 1: Deny A Complex/Simple Asymmetry	39
2.5 Response 2: Continuity Relations Matter	53
2.6 Conclusion	73
3 Grounds For Concern	77
3.1 The Importance Of The More Particular Facts	78
3.2 Self-Constitution Accounts	102
3.3 Conclusion	123
4 Transplant Intuitions	127
4.1 Locke’s <i>Prince And The Cobbler</i>	129
4.2 Williams’ <i>Cases</i>	131
4.3 Transplant Intuitions Are Off-Track	136
4.4 Debunking Transplant Intuitions	149
4.5 Debunking And Objectivism About Personal Identity	152
4.6 Debunking And Skepticism About Personal Identity	160
4.7 Conclusion	164
5 Against The Complex Versus Simple Distinction	167
5.1 Parfit’s Definitions	169
5.2 Gasser And Stefan	174
5.3 Noonan	180

CONTENTS – COMPACT

5.4	Two Purposes Of A Distinction	187
5.5	Conclusion	192
6	Replacing The Complex Versus Simple Distinction	195
6.1	Ontological Dependence	197
6.2	Fundamentalism About Persons	206
6.3	Essentialism And Haecceitism	213
6.4	Conclusion	219
7	Personal Identity And Transformative Experiences	223
7.1	Transformative Experience And Revelation	225
7.2	Paul and Shupe On Cochlear Implants	230
7.3	Parent-Focus	235
7.4	Harms And Other Consequences	242
7.5	Surrogate Decision-Making	248
7.6	A Proposal On Cochlear Implants	255
7.7	Conclusion	261
8	Transformative Choice And Informed Consent	263
8.1	Protection	266
8.2	Autonomy	267
8.3	Self-Realization	275
8.4	Conclusion	277
	Conclusion	279
	Bibliography	287

Contents – Detailed

Introduction	1
1 The Complex Versus Simple Distinction	9
1.1 Positive Reasons	11
1.1.1 In Favour Of The Simple View	11
1.1.2 In Favour Of The Complex View	15
1.2 Negative Reasons	17
1.2.1 Against The Simple View	17
1.2.2 Against The Complex View	20
1.3 Conclusion	24
2 The Extreme Claim	27
2.1 Practical Objections	29
2.2 The Extreme And The Moderate Claim	33
2.3 A Reductio Against The Complex View	38
2.4 Response 1: Deny A Complex/Simple Asymmetry	39
2.4.1 The Simple View And Concern	39
2.4.2 Sources Of The Reasonableness of Concern	43
2.4.3 Dependence Between Personal Identity And Concern	50
2.5 Response 2: Continuity Relations Matter	53
2.5.1 Against Numerical Distinctness: Persistence	54
2.5.2 Quasi-Anticipation	59
2.5.3 Insights From Fission	67
2.5.4 Ordinary Survival In The Complex View	70
2.6 Conclusion	73
3 Grounds For Concern	77
3.1 The Importance Of The More Particular Facts	78
3.1.1 Whiting On Friends And Future Selves	79
3.1.2 Brink On The Separateness Of Persons	86
3.1.3 Reflections: The Constitutive Role Of Attitudes	94
3.2 Self-Constitution Accounts	102
3.2.1 Korsgaard On The Unity Of Agency	103

CONTENTS – DETAILED

3.2.2	Schechtman On Narrative Identity	109
3.2.3	Reflections: Self-Constitution And Psychological Continuity	117
3.3	Conclusion	123
4	Transplant Intuitions	127
4.1	Locke’s <i>Prince And The Cobbler</i>	129
4.2	Williams’ <i>Cases</i>	131
4.3	Transplant Intuitions Are Off-Track	136
4.3.1	Quality Of Evidence: <i>Case 1</i>	136
4.3.2	Quality Of Evidence: <i>Case 2</i>	141
4.3.3	Quasi-Attitudes	143
4.3.4	Divergence	148
4.4	Debunking Transplant Intuitions	149
4.5	Debunking And Objectivism About Personal Identity . .	152
4.6	Debunking And Skepticism About Personal Identity . . .	160
4.7	Conclusion	164
5	Against The Complex Versus Simple Distinction	167
5.1	Parfit’s Definitions	169
5.1.1	Introduction of Complex/Simple And Reductionism/- Non-Reductionism	169
5.1.2	The Relation Between Complex/Simple And Reduct- ionism/Non-reductionism	170
5.1.3	Parfit And Misclassifications	172
5.2	Gasser And Stefan	174
5.2.1	Analysability	176
5.2.2	Consist-In Claims	177
5.2.3	Necessary And Sufficient Conditions	177
5.2.4	Circularity/Informativeness	178
5.3	Noonan	180
5.3.1	The Proposal	180
5.3.2	Refinement	182
5.3.3	Being Identity-Involving	185
5.4	Two Purposes Of A Distinction	187
5.5	Conclusion	192
6	Replacing The Complex Versus Simple Distinction	195
6.1	Ontological Dependence	197
6.2	Fundamentalism About Persons	206
6.3	Essentialism And Haecceitism	213
6.4	Conclusion	219

CONTENTS – DETAILED

7	Personal Identity And Transformative Experiences	223
7.1	Transformative Experience And Revelation	225
7.2	Paul and Shupe On Cochlear Implants	230
7.3	Parent-Focus	235
7.3.1	Strategy 1: Parent-Focus Is Not Implausible . . .	236
7.3.2	Strategy 2: The Revelation Approach Does Not Apply	240
7.4	Harms And Other Consequences	242
7.4.1	Strategy 3: Enrich Revelatory Values	243
7.4.2	Strategy 4: The Revelation Approach Does Not Apply	245
7.5	Surrogate Decision-Making	248
7.5.1	Best Interests	249
7.5.2	Substitute Judgement	251
7.5.3	Strategy 5: The Revelation Approach Is Innocent	253
7.6	A Proposal On Cochlear Implants	255
7.7	Conclusion	261
8	Transformative Choice And Informed Consent	263
8.1	Protection	266
8.2	Autonomy	267
8.2.1	Dworkin	268
8.2.2	Faden And Beauchamp	269
8.2.3	Savulescu	270
8.2.4	DeGrazia	272
8.3	Self-Realization	275
8.4	Conclusion	277
	Conclusion	279
	Bibliography	287

Introduction

You, I, and the people around us are persons. We all existed yesterday, and will exist tomorrow, remaining one and the same. Personal identity over time seems ubiquitous.

That each of us remains one and the same person over time does not only seem obvious, but also appears to play a crucial role for many practical concerns. Perry illustrates:

“Most of us have a special and intense interest in what will happen to us. You learn that someone will be run over by a truck tomorrow; you are saddened, feel pity, and think reflectively about the frailty of life; one bit of information is added, that the someone is you, and a whole new set of emotions rise in your breast.”¹

In memory, pride, and regret, we take it for granted that it was the very same person as our present self who carried out an earlier action. In concern, anticipation, investing, and planning, we presuppose that the very same person as the present, forward-looking self will experience the consequences of its current and past efforts and actions. In holding someone morally responsible, we presuppose that it was the very same person whom we are now blaming who carried out the wrongful act. Some people anticipate an afterlife, and in doing so they seem to presuppose that they will be identical to an entity on the other side.²

What does it take for these presuppositions about personal identity to be true? First of all, let us be clear about the topic. There are a number of distinct questions about personal identity that are not always kept apart.³ I will be concerned primarily—and unless noted otherwise—with what I call

¹Perry 1976, p. 67.

²For further examples, see D. Shoemaker 2007, pp. 317-8.

³See also Olson 2017, Ch. 1, although my framing differs slightly from his.

The Diachronic Metaphysical Question: “If a person x exists at one time and something y exists at another time, under what possible circumstances is it the case that x is y ?”⁴

In order to address this question, I will occasionally—especially in (chapter 4)—turn to a related but distinct question:

The Diachronic Epistemic Question: If a person x exists at one time and something y exists at another time, what counts as evidence that x is y ?

We can ask further questions about persons:

The Synchronic Metaphysical Questions: In virtue of what is an entity a person, and in virtue of what is she a particular person?

The Synchronic Epistemic Questions: What counts as evidence for something’s being a person, and for being a particular person rather than some other person?

Colloquially, we sometimes speak of a person’s identity. In such cases, we appear to refer to a set of properties that are thought to be distinctive of a particular person, and make her the person she is. This way of speaking can be understood as an attempt to address one of the synchronic questions. In the following, the synchronic questions will pop up occasionally, for example in (3.2), (chapter 7), and (chapter 8).

Currently, the most popular candidates for answering the diachronic metaphysical question are psychological continuity theory and animalism. Psychological continuity theory claims that person x at time t_1 is identical to y at t_2 if and only if x is psychologically continuous with y . Animalism claims that we are identical to animals, and hence that person x is identical to y if and only if they are the same human animal.

⁴[Ibid.](#), sect. 2.

There are actual and possible cases in which these theories give different verdicts. Suppose y is in a permanent vegetative state. Animalists would not hesitate to deem x and y identical if they are the same human animal. Psychological continuity theorists would disagree; x and y are distinct, given that y is psychologically discontinuous with x by virtue of—let us suppose—having no psychological life left at all.

Or consider a futuristic scenario in which q , who has a human body but formerly had no cerebrum, is the recipient of a transplant of p 's cerebrum.⁵ The procedure goes well and preserves psychological continuity. Psychological continuity theory thus would judge that q is p . Animalists would deny this on the grounds that they are not the same human animal.

A further view which has been popular in the history of philosophy, but is endorsed less often nowadays, is that x is y if x has, or is, the same soul, ego, or monad as y . The theories differ in whether and how closely these entities are tied to the continued existence of a particular body and/or a particular mental life. However, some versions, such as what Parfit⁶ calls the Featureless Cartesian View, allow for personal identity to completely float free of any observable facts.

The practical concerns⁷ mentioned at the outset appear to presup-

⁵Olson (1997, pp. 44-6) argues that it is crucial that only the cerebrum, not the whole brain, is transplanted. The latter includes the brain-stem, which is essential to the human animal's biological functioning. Following Van Inwagen (1990, pp. 172-181), it is an option to see the whole brain as a pared down human animal. Psychological continuity theory and animalism would then *agree* that q is p , but for different reasons: the former because psychological continuity has been preserved, the latter because they are the same human animal.

⁶Parfit 1984, p. 228.

⁷Like Olson (1997, ch. 3), D. Shoemaker (2007), and others, I use 'practical concerns' as an umbrella term for practices like concern, anticipation, compensation, moral responsibility, pride, regret, etc. As will become clear, this slides over certain

pose personal identity. The question arises what this means for our theorizing about the persistence of persons. Should our theory choice be constrained in some way by how our practices appear to make sense of personal identity? That is, do our practices have implications for what can count as the right theory of personal identity? Or is it the other way round: should we identify the right theory of personal identity on purely theoretical grounds, and then come to care, hold responsible, anticipate, compensate, etc. in accordance with this theory? Or is neither of these dependences appropriate? Is maybe the starting point that practical concerns operate with a notion of personal identity misguided?

As we will see in the following, authors have answered each of these questions in the affirmative. I myself am convinced that the interdependence between personal identity and the mentioned practical concerns is overstated. In paradigmatic places where philosophers and common sense suggest that personal identity constrains how we should reason and care, or vice versa, I argue that the two spheres are in fact neutral to each other. I will argue for these claims in the course of the following eight chapters.

Chapter 1: The Complex Versus Simple Distinction. Debates about personal identity are often framed in terms of the distinction between the complex and the simple view. I describe some reasons authors give in favour of their preferred view and against its competitor.

differences amongst these activities. For example, some activities might differ from others in how they are justified, and how their justification makes reference to personal identity. For most of the following, I will discuss the specific example of future-directed concern. I will attempt to highlight if and when other practical activities differ in important ways.

Chapter 2: The Extreme Claim. According to the extreme claim, the complex view fails to provide reasons for future-directed concern. Some authors take the extreme claim as a premise in a *reductio* argument against the complex view. I describe two broad strategies to counter the *reductio*. First, by denying that there is an asymmetry between the complex and the simple view in terms of how well they justify concern. Second, by arguing that the relations of continuity and connectedness to which the complex view refers do not diminish the value of ordinary survival.

Chapter 3: Grounds For Concern. I turn to the positive counterpart of the claim that the complex view does not diminish the value of ordinary survival. I examine proposals according to which what *does* matter in ordinary survival might be related to personal identity; but it is something else. We maintain forward-looking attitudes and projects not because someone will be *us*, but because we relate to future selves in other, more important ways. According to the first class of positions, the relations of continuity and connectedness to which the complex view refers provide reasons for future-directed concern. According to the second class of positions, we need to look beyond the numerical identity of persons over time and/or what it consists in.

Chapter 4: Transplant Intuitions. I will understand transplant intuitions as judgements about where a person goes if psychological and bodily continuity come apart. Coherence with transplant intuitions is often taken to be a point in favour of a theory of personal identity, and incoherence with these intuitions as a disadvantage. I argue that transplant intuitions are unsuitable for tracking truths about personal

identity. By taking our person-directed attitudes as evidence for where a person goes, transplant intuitions are contaminated with practical concerns whose relevance for personal identity is far from straightforward. I conclude that transplant intuitions provide only circular evidence, seem entirely uninformative if quasi-attitudes are possible, and diverge inexplicably across cases.

Chapter 5: Against The Complex Versus Simple Distinction. I examine three proposals on the difference between the complex and the simple view: Parfit's original introduction of the distinction, Gasser and Stefans definition, and Noonan's recent proposal. I argue that the first two classify the paradigm cases of simplicity as complex, while Noonan's proposal makes simplicity and complexity turn on features whose relevance for the distinction is questionable. Given these difficulties, I examine why we should be interested in whether a position is complex or simple. I describe two purposes of having a distinction, and show that extant accounts of the complex versus simple distinction fail to serve these. I argue that unless we find a satisfying account of the difference between complex and simple positions, we should not frame discourses on personal identity in these terms.⁸

Chapter 6: Replacing The Complex Versus Simple Distinction. After having criticized the complex versus simple distinction, I owe an account of how else we should frame discourses about personal identity. I argue that we should disambiguate and debate more refined questions and proposals. I sketch and discuss three candidate notions for expressing and debating claims about personal identity: ontological

⁸This material appears in Hummel 2017.

dependence, fundamentality, and essentialism.

Chapter 7: Personal Identity and Transformative Experience. I explain Paul's notion of transformative experience and her revelation approach for making transformative choices. I defend the revelation approach against objections by Shupe, according to which the revelation approach is inappropriate if decision-maker and transforming individual are numerically distinct.

Chapter 8: Transformative Choice and Informed Consent. I argue that contrary to what Paul appears to suggest, the fact that patients and research subjects are unable to projectively imagine transformative diseases and treatments does not diminish the importance of informed consent procedures. I characterize three purposes of informed consent, and argue that an individual's failure to projectively imagine an outcome does not prevent informed consent from fulfilling these purposes.

In the following, I will often refer to the complex versus simple distinction. In such cases, I am not contradicting my own advice (chapters 5, 6) that debates involving personal identity should not be framed in terms of this distinction. Instead, I am making a concession. I grant that there is something to the distinction. But I demonstrate that even then, the conclusions which authors have drawn on the basis of this distinction do not follow.

Chapter 1

The Complex Versus Simple Distinction

A common distinction in the contemporary debate about personal identity is between the complex and the simple view. Parfit defines:

“According to the Complex View, the fact of personal identity over time just consists in the holding of certain other facts. It consists in various kinds of psychological continuity, of memory, character, intention, and the like, which in turn rest upon bodily continuity. According to the Simple View, personal identity does not just consist in these continuities, but is a quite separate ‘further fact.’”⁹

In the following, I describe a selection of reasons why authors believe that this distinction is important. The purpose of this chapter is merely to provide an understanding of the complex versus simple distinction in order to set the stage for the following chapters. I do not intend provide an exhaustive overview on the reasons that motivate authors to take sides in the debate. Moreover, as mentioned in the Introduction, I will eventually argue that the complex versus simple distinction is confused, and should not be used to frame discourses on personal identity (chapter 5). Until then, I grant that there is something to the complex versus simple distinction. In the following, I describe what authors think it is.

Since the complex versus simple distinction was first drawn by Parfit, earlier authors do not mention it. But paradigm cases of complexity and simplicity precede the introduction of the distinction. I will focus on the reasons why these authors think their positions are correct, and why they deem paradigmatic instances of the competing view mistaken.

⁹Parfit 1982, p. 227.

1.1 Positive Reasons

I begin by describing reasons put forward *in favour* of either side.

1.1.1 In Favour Of The Simple View

Insofar as the simple view accepts accounts of personal identity, they typically refer to the persistence of souls, Cartesian egos, or monads. One *prima facie* reason in favour of these pictures is their popularity and dominance in the history of philosophy. Martin writes:

“[f]rom Plato until John Locke personal identity was explained in the West primarily by appeal to the notion of a spiritual substance or soul.”¹⁰

Especially for Christian thinkers, one important attraction of such a position arises from the features typically ascribed to souls: “immaterial, indivisible, and hence naturally immortal.”¹¹ Souls and their persistence were—and for many remain—crucial in explanations of how persons can anticipate an afterlife.

Simple positions thus respect deeply held intuitions about what a person can survive. In particular, there is nothing self-contradictory in supposing that a person can survive the death of her body, or even a sudden and radical change of her psychology. Swinburne writes:

“[m]any religions have taken seriously stories of persons passing through the waters of Lethe (a river whose waters made a person forget all his previous life) and then acquiring a new body. Others who have heard these stories may not have believed them true; but they have usually claimed to understand them, and (unless

¹⁰Martin 1998, p. 1.

¹¹*Ibid.*, p. 3.

influenced by philosophical dogma) have not suspected them of involving contradiction.”¹²

Similarly, Chisholm maintains that it is intelligible to suppose that survival is compatible with the most drastic changes. Referring to (and twisting) a hypothetical case suggested by Leibniz,¹³ Chisholm writes:

“Since I *am* concerned about my future welfare, about the welfare of that *x* such that *x* is identical with me, and presumably will continue to have this concern, it will follow that, if I should learn that *x* will some day become King of China, then, even though I also learn that *x* will then forget everything that *x* had previously been, I will be as much concerned about the life that *x* enjoys while King of China as I am about my own; for, given our suppositions, *x*’s welfare will be my own.”¹⁴

According to these authors, it is a point in favour of the simple view that it does not rule out these intelligible and conceivable scenarios. They suggest that our survival is entirely independent from what Parfit calls the more particular facts. Madell is convinced that

“consideration of what we can coherently imagine for ourselves must lead us to reject the empiricist’s claim to provide a sufficient analysis of personal identity in terms of any sort of observable connection between experiences”¹⁵,

and also that

“the claim that objective connections between experiences are a necessary condition only of personal identity through time is totally untenable”¹⁶.

One motivation for the idea that personal identity can float free from other facts and allow a person to survive even the most radical bodily

¹²Swinburne 1984, p. 25.

¹³Leibniz 1686, sect. 34.

¹⁴Chisholm 1969, p. 138.

¹⁵Madell 1981, p. 104.

¹⁶*Ibid.*, p. 107.

and psychological discontinuities is its plausibility from a first-person point of view. Ninan writes:

“[h]ere’s something I like about the Simple View: it seems true when I consider my own persistence through time from the inside or from the first-person point of view. I’m not alone in finding this aspect of the Simple View appealing. A recurring theme in the literature on our persistence is that, when one considers ones own persistence from the first-person point of view, it seems as though ones persistence is a basic and fundamental fact, one that is metaphysically independent of facts about physical and psychological continuity.”¹⁷

Similarly, Johnston remarks that consciousness presents itself as constant through both psychological and bodily changes:

“we can imagine many sorts of cases that seem to involve one’s ceasing to be associated with a particular human body and human personality. These cases are particularly compelling when imagined ‘from the inside.’ So I am to imagine myself undergoing a radical change in my form, e.g., a change like that undergone by Franz Kafka’s beetleman, and perhaps concurrently a wild change in my psychology. There seems to be nothing internally incoherent about such imaginings. [...] [I]t seems that such imaginings are not ruled out by and so are consistent with our concept of a person.”¹⁸

Johnston’s point here is that the persistence of a person appears compatible with the wildest discontinuities.

Another, complementary intuition which the simple view respects can be illustrated by means of fission, i.e. the symmetrical branching of bodily and/or psychological continuity. Suppose that before undergoing fission, the fissioner may decide which fission product gets tortured, and which one receives a reward. Swinburne writes:

¹⁷Ninan 2009, pp. 429-30.

¹⁸Johnston 1987, p. 70. He goes on to criticize such imaginings as “idle” (*Ibid.*).

“We can make sense of the supposition that the victim makes the wrong choice, and has the experience of being tortured and not the experience of being rewarded; or the right choice, and has the experience of being rewarded and not the experience of being tortured. A mere philosophical analysis of the concept of personal identity cannot tell you which experiences will be yours tomorrow. To use Bernard Williams’s telling word, any choice would be a ‘risk’.”¹⁹

Besides Johnston’s point that personal identity appears compatible with the wildest discontinuities, fission illustrates that even if we hold bodily or psychological continuities *fixed*, this seems compatible with *different* identity facts.

Swinburne’s positive view is that the fissioner is identical to the one offshoot that receives her soul. This leads to yet another feature of the simple view: it is *determinate* which offshoot has the fissioner’s soul even if both offshoots are similarly continuous with the fissioner.²⁰ Again, this is plausible from the first-person point of view. As Blackburn illustrates, imagine one fission product awakes in pain in a noisy red room, and the other awakes painlessly in a silent green room. From the inside,

“there seem in prospect to be only three possibilities: at that time either I will experience pain, noise and red, or lack of pain, noise and green, or I will not awake at all. There seems, in advance, to be no other relevant possibility.”²¹

The simple view fits this data well. In Swinburne’s terms, the soul either goes with one fission product, or the other, or ceases to exist.

¹⁹Swinburne 1984, p. 18.

²⁰*Ibid.*, pp. 18-20.

²¹Blackburn 1997, p. 181. He calls this the *unity reaction*, and actually denies that it reflects the plausibility of a Cartesian, i.e. simple picture. Instead, he suggests that broadly Kantian considerations are necessary to understand the reaction.

Chisholm gives a similar judgement on the grounds that persons are mereologically simple—a claim which not all simple positions accept.²² Chisholm believes that the bearer of my consciousness is a particular atom in my brain.²³ In his case, this belief is tied to several distinctive claims, such as:²⁴ only mereologically constant objects are identical over time *in the strict philosophical sense* (see 1.2.2); mereologically variable entities are merely identical in *the loose and popular sense*, and are logical constructions out of the former; mereologically variable objects have their properties that are “rooted only in the times at which they are had”²⁵—roughly: intrinsic properties—in virtue of the mereologically constant objects having these properties; mental properties such as “that I now hope for rain”²⁶ are had by myself and not borrowed from other things.

Some think that it is a central, pre-theoretical intuition of ours that conscious beings are mereologically simple.²⁷ Chisholm’s simple position respects this intuition.

1.1.2 In Favour Of The Complex View

Parfit endorses the complex view on the basis of considering a range of thought experiments. One central case is the *combined spectrum*: Parfit’s body and psychology get gradually altered and replaced until a perfect replica of Greta Garbo results. Parfit distinguishes three reactions:

²²Swinburne 1984, p. 21.

²³Chisholm 1986, p. 75.

²⁴Chisholm 1976, ch. 3.

²⁵*Ibid.*, p. 104.

²⁶*Ibid.*

²⁷According to Barnett (2008, 2010), the idea that consciousness requires mereological simplicity is part of our naive conception of conscious subjects. In contrast, Madden (2015) argues that this conception makes claims not about mereology, but topology: conscious subjects must be topologically integrated, not scattered.

“accepting a Reductionist reply, believing that there must be some sharp borderline, and believing that the resulting person would in every case be me.”²⁸

It is implausible to hold that Parfit remains himself throughout the procedure, especially at the end of the case where perfect resemblance to Greta Garbo has been brought about. But it is also absurd to believe that there is a point where, through one tiny and seemingly trivial bit of extra change, a threshold of psychological and/or physical discontinuity is exceeded such that Parfit does not survive.

“It is hard to believe that the difference between life and death could just consist in any of the very small differences”²⁹.

Not only is it implausible to believe that a tiny difference determines whether Parfit survives. If there was such a borderline, it seems we could never find out where it lays and when it is exceeded. It is hard to believe that

“there must be such a sharp borderline, somewhere in the Spectrum, though we could never have any evidence where the borderline would be.”³⁰

It is more plausible to accept the complex view according to which personal identity consists in bodily and/or psychological facts. Parfit exists at the outset and does not exist at the end of the spectrum. But it is not the case that for each point in the spectrum, Parfit either determinately survives or determinately does not survive. There is a range where it is indeterminate whether he still exists. In these cases,

²⁸Parfit 1984, p. 238.

²⁹*Ibid.*, p. 239.

³⁰*Ibid.*

“though we know the answer to every other question, we have no idea how to answer a question about personal identity.”³¹

The latter question is *empty*: different answers do not denote different possibilities, just different descriptions of one and the same possibility.³²

Further support for the complex view comes from empirical observations. We have not yet seen facts of personal identity in the absence of physical, biological, or psychological facts. This provides inductive evidence for the idea that person facts consist in more particular facts, and that they do so as a matter of nomological or some other kind of necessity. This fits neatly with broadly naturalistic, physicalist pictures—at least provided that psychological or mental facts, if relevant to personal identity, are not external to natural science. The complex view thereby “cohere[s] with the rest of what we know about the world.”³³

1.2 Negative Reasons

Each of the views offers benefits. But there are also costs. I now describe reasons *against* either side.

1.2.1 Against The Simple View

While the complex view appears to cohere well with the rest of what we know about the world, authors note that the simple view lacks empirical support. Parfit thinks that we might have had evidence for the simple view. He provides the hypothetical case of a Japanese woman who reports

³¹Parfit 1971, p. 3.

³²Parfit 1995, p. 23.

³³S. Shoemaker 1984, p. 71.

having lived a past life as a Celtic warrior. If we encountered such reports and could confirm their accuracy, we might have

“to abandon the belief that the carrier of memory is the brain. [...] We might have to assume that there is some purely mental entity, which was in some way involved in the life of the Celtic warrior, and is now in some way involved in the life of the Japanese woman [...]. A Cartesian ego is just such an entity.”³⁴

However, the problem is that

“we have no evidence to believe that psychological continuity depends chiefly, not on the continuity of the brain, but on the continuity of some other entity”³⁵.

Similar points are being made by Johnston. With regards to soul views, he notes that

“the existence of the soul and consequently the nature of personal identity are empirical questions, to be settled by our total evidence.”³⁶

And the mental properties that are distinctive of persons

“seem to have definite brain functions as their conditions sine qua non. [...] Even in cases of recovery from the specific cognitive losses produced by local brain damage, there is, significantly, no reported phenomenology of memories of an intact thinking soul being ‘locked inside’ an inept, because damaged, brain and body. The thoughts and mental capacities were just not there, it seems.”³⁷

In principle, out-of-body or near-death experiences appear to track facts of personal identity that are independent from more particular facts, in Johnston’s case facts about the brain. But Johnston complains that there

³⁴Parfit 1984, p. 227.

³⁵*Ibid.*, p. 228.

³⁶Johnston 2010, p. 135.

³⁷*Ibid.*, p. 130.

is no actual evidence that such experiences ever resulted from instances in which brain facts were *absent*. Instead, as things stand, these reports are to be treated on par with reports of consumers of hallucinogenic substances that bring about brain states of having out-of-body experiences. Nothing about these reports establishes that their content is not illusory.

The fact that there are no actual cases where personal identity is secured in the absence of psychological and/or bodily relations brings the simple view under pressure. The question arises why we should believe that personal identity is a further fact, when we have never seen an instance of personal identity in the absence of the more particular facts.

The simple view faces further challenges. For example, it tends to be a merely negative view. It rejects the complex view, but does not always provide a sufficiently detailed positive picture that could replace the view deemed false. As one consequence, it is not obvious why the simple view should be attractive. On the contrary, the independence of personal identity from more particular facts seems compatible with two worlds being indiscernible at the micro-level, but differing with regards to facts about personal identity. This is how Zimmermann paraphrases positions endorsed by Lowe³⁸ and Merricks³⁹. Zimmerman deems these “truly heroic”⁴⁰, incredible, and unmotivated. He labels such views *identity mysticism*. These positions seem dubious because the connection between micro-level facts and personal identity looks mysterious, magical, or even unintelligible.

³⁸Lowe 1996.

³⁹Merricks 1998.

⁴⁰Zimmerman 1998, p. 295.

1.2.2 Against The Complex View

The complex view takes personal identity to consist in more particular facts. Proponents of the simple view object that these more particular facts presuppose rather than constitute personal identity. For example, Butler objects against Locke’s psychological account that it is

“self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute personal identity, any more than knowledge in any other case, can constitute truth, which it presupposes”⁴¹.

One way to read this worry is that as a matter of conceptual necessity, consciousness- and/or memory-continuity require that the connected person stages are (part of) the same person. This means that in the Lockean picture of personal identity,

“the concept allegedly being defined is illicitly employed in the formulation of the defining definition.”⁴²

In response, S. Shoemaker suggests the notion of quasi-memory, which is just like memory, but does not presuppose personal identity as a matter of conceptual necessity.⁴³ It can relate mental states of numerically distinct persons. An account of personal identity in terms of quasi-memory thus has the potential to avoid this particular circularity charge.

According to a different objection from Butler and Reid, the complex view fails to give an account of personal *identity* over time. Butler argues that an object which loses and gains parts over time, such as a tree, can remain the same “as to all the purposes of property and uses of common

⁴¹Butler 1736, p. 100.

⁴²S. Shoemaker 1970, p. 281

⁴³*Ibid.*, p. 271.

life”⁴⁴. Reid agrees that “it retains the same name, and is considered as the same thing”⁴⁵. However, because of the change in parts, the tree’s similarity “is rather something which, for the convenience of speech, we call identity”⁴⁶.

In contrast, persons remain the same “in the strict philosophical sense of the word”⁴⁷. The latter “cannot subsist with diversity of substance”⁴⁸, where *substance* seems to refer to what a thing is made up of, or what it consists in.⁴⁹ As Reid puts it:

“The identity of a person is perfect identity: wherever it is real, it admits of no degrees; and it is impossible that a person should be in part the same, and in part different; because a person is a *monad*, and is not divisible into parts.”⁵⁰

Lockean positions reduce personal identity to relations amongst conscious states. Butler worries that consciousness over time does not remain one and the same thing. It consists in a sequence of distinct, successive conscious states, e.g., perceptions at different times. Proponents of a Lockean picture thus fail to capture identity in the strict philosophical sense. They must think that

“personality is not a permanent, but a transient thing: that it lives and dies, begins and ends, continually: that no one can any more remain one and the same person two moments together, than two successive moments can be one and the same moment: that our substance is indeed continually changing; but whether this be so or not, is, it seems, nothing to the purpose; since it is not

⁴⁴Butler 1736, pp. 99-100.

⁴⁵Reid 1785b, p. 112.

⁴⁶*Ibid.*

⁴⁷Butler 1736, p. 100.

⁴⁸*Ibid.*

⁴⁹S. Shoemaker (1984, pp. 73-4) argues that this is a specific and rather unhelpful understanding of *substance*. There is an alternative, Aristotelian understanding according to which a mereologically changing tree does remain the same substance.

⁵⁰Reid 1785b, p. 111.

substance, but consciousness alone, which constitutes personality: which consciousness being successive, cannot be the same in any two moments, nor consequently the personality constituted by it.”⁵¹

According to the complaint, the Lockean account is not really an account of *identity* over time. Conscious states at different times are distinct. The persons that supposedly consist in these states thus must remain distinct, too. Personal *identity* falls out of the picture. I call this the *numerical distinctness charge*.

Butler and Reid are convinced that the Lockean picture would have bizarre practical consequences:

“it must follow, that it is a fallacy upon ourselves, to charge our present selves with any thing we did, or to imagine our present selves interested in any thing which befell us yesterday; or that our present self will be interested in what will befall us tomorrow: since our present self is not, in reality, the same with the self of yesterday, but another like self or person coming in its room, and mistaken for it; to which another self will succeed tomorrow. This, I say, must follow: for if the self or person of today, and that of tomorrow, are not the same, but only like persons; the person of today is really no more interested in what will befall the person of to morrow, than in what will befall any other person.”⁵²

Because of the numerical distinctness between present and future person stages, attitudes of concern and ascriptions of responsibility for past or future events become unreasonable.⁵³

A final worry about the complex view relates to indeterminacy. The more particular facts can come in degrees. For example, direct quasi-memory connections between person stages can be few or many, and the resulting degree of psychological connectedness can be low or high. There

⁵¹Butler 1736, p. 102.

⁵²*Ibid.*

⁵³I will discuss these claims in (chapter 2).

are thus conceivable instances, such as the ‘range’ discussed earlier in Parfit’s *Spectrum*, in which an individual neither determinately survives nor determinately ceases to exist. Firstly, the possibility of such cases is surprising and counterintuitive.

“When it is applied to ourselves, this Reductionist claim is hard to believe. In such imagined cases [where personal identity is indeterminate; P.H.], something unusual is about to happen. But most are inclined that, in any conceivable case, the questions ‘Am I about to die?’ must have an answer. And we are inclined to believe that this answer must be either, and quite simply, Yes or No. Any future person must be either me, or someone else. These beliefs I call the view that *our identity must be determinate*.”⁵⁴

Secondly, it is quite unclear which attitudes we should have towards such cases. As Williams puts it, the prospect of indeterminacy “seems to have no comprehensible representation in my expectations and the emotions that go with them.”⁵⁵ If it is indeterminate whether a present self is identical to a future self who will get tortured,

“fear [...] seems neither appropriate, nor inappropriate, nor appropriately equivocal. Relatedly, the subject has an incurable difficulty about how he may think about [situation] *S*. If he engages in projective imaginative thinking (about how it will be for him), he implicitly answers the necessarily unanswerable question; if he thinks that he cannot engage in such thinking, it looks very much as if he also answers it, though in the opposite direction.”⁵⁶

These reasons against a view that allows for indeterminacy need not be decisive. But they point to puzzles that loom within the picture.

⁵⁴Parfit 1984, p. 214.

⁵⁵Williams 1970, p. 175.

⁵⁶*Ibid.*, pp. 176-7.

1.3 Conclusion

The simple view is somewhat intuitive. From the inside, our survival seems compatible with a wide range of bodily and mental changes. And for one and the same set of facts and relations of bodily and mental continuity and connectedness, it might seem an open question whether the scenario contains *us*. Complex theorists accept that we might have had good evidence for the simple view. But as a matter of fact, it is not well-supported by further evidence beyond these intuitions and seemings from the inside.

The complex view too has intuitive aspects. In Parfit's *Spectrum*, it seems plausible that Parfit does not remain himself throughout the procedure, and that there is no sharp cut-off point where the tiniest change makes all the difference about his survival. It is somewhat attractive to believe that Parfit's survival depends on the more particular facts that are gradually altered by the procedure. This idea also coheres well with a broadly naturalistic world view, given that personal identity is taken to be nothing over and above bodily and/or mental facts and relations. But the complex view also faces questions. It is quite unclear which attitude we should take towards cases where personal identity is indeterminate. Moreover, if personal identity consists in facts and relations which by themselves seem far less important than our survival, how can it make sense to care about survival, to be concerned about the past and future, to ascribe moral responsibility, and to engage in other person-related practices? This issue might seem particularly pressing in view of the insight that what Parfit calls the more particular facts at different times

are *distinct*. For example, my present conscious states are plausibly understood to be numerically distinct from my conscious states tomorrow. Does this not give a sense in which I fail to remain one and the same person over time, and in which my engagement in person-related practices is based on a mistake? I will discuss these and other questions in the following.

Chapter 2

The Extreme Claim

Most people maintain some future-directed concern. We plan, set goals, prepare, imagine, compare scenarios, make choices, look forward to future joy and fearfully anticipate pain, and hope that things work out. Even people who live in the moment and do not worry about tomorrow tend to care about more than the present; they just do not extend their concerns as far into the future as others. One assumption on which these activities rest is that it will be *us* who experience the consequences, benefit from success, take part in the happiness of others, and suffer from accidents or mistakes. When caring in these ways, we make assumptions—if only minimal ones—about our survival into the future.

Unfortunately, assumptions can turn out to be false, and this can affect our investment into activities which presuppose them. If I was told that I will not exist tomorrow, or that I will exist but not in a way that matters to me, I would begin to live differently today. I might seize the day, prioritize efforts, and refrain from working towards goals that have become unreachable. I need not stop thinking and caring about the future. But since I myself will not be part of it, my current perspective on it has changed.

Some philosophers think that we need not receive particularly bad news to be forced to change our attitudes in such ways. If we started believing the complex view, no person could take herself to survive into the future, at least not in a sense that matters. Concern for our futures would become a deeply misguided attitude. If all that being and remaining *you* involves is the obtainment of relations of continuity and connectedness, there is no reason to care about what happens tomorrow, or at least no more reason to care about what happens to *you* than to

care about more general matters.

These prospects are troubling. The complex view might be true—does this mean that our future-directed concern is unjustified, and that we should care more about the present? On the other hand, our current attitudes seem clearly justified and so natural that any theory which conflicts with them looks dubious—are we entitled to criticize or dismiss the complex view on the basis of our pre-theoretic concern about our futures?

In the following, I will discuss these questions, and eventually answer both of them in the negative. I begin by describing aspects in which the complex view is supposed to undermine future-directed concern (2.1), before distinguishing two more general claims in view of these findings (2.2). I then sketch an argument for how these practical objections supposedly lead to a *reductio* of the complex view (2.3), before characterizing several promising variants of two broad responses (2.4, 2.5) to the idea that the complex view has implausible practical implications. In the end, everything will be fine. We are justified in caring about the future even if the complex view is true. And when choosing between the complex and the simple view, we can set our everyday concerns aside, and, if we prefer, endorse the best option on purely theoretical grounds.

2.1 Practical Objections

The complex view has been accused of having absurd practical consequences. According to the critics, these have at least two sources.⁵⁷

⁵⁷Whiting 1986, pp. 549-52.

The first source is related to numerical distinctness. As mentioned (1.2.2), Butler complains that for Locke, person stages at different times are related by similarity of consciousness, but remain otherwise *distinct*. Thus, no account of *identity* has been offered.

“[I]t must follow, that it is a fallacy upon ourselves, to charge our present selves with any thing we did, or to imagine our present selves interested in any thing which befell us yesterday, or that our present self will be interested in what will befall us tomorrow; since our present self is not, in reality, the same with the self of yesterday, but another like self or person coming in its room, and mistaken for it; to which another self will succeed tomorrow.”⁵⁸

Since in Locke’s picture, present and future self are numerically distinct, the former has no reason to care about the latter. They are similar selves, but not the same self. Butler thinks that holding accountable and caring about the future are attitudes that presuppose numerical identity. Thus, the Lockean cannot avoid making

“the inquiry concerning a future life of no consequence at all to us, the persons who are making it”⁵⁹

Locke’s position that personal identity consists in sameness of consciousness inspires psychological continuity theory. Lewis writes:

“I find what I mostly want in wanting survival is that my mental life should flow on. My present experiences, thoughts, beliefs, desires, and traits of character should have appropriate future successors [...] These successive states should be interconnected in two ways. First, by bonds of similarity. Second, by bonds of lawful causal dependence.”⁶⁰

Such Neo-Lockean positions have received similar criticisms. For exam-

⁵⁸Butler 1736, p. 102.

⁵⁹*Ibid.*, p. 99.

⁶⁰Lewis 1976, p. 17.

ple, Schechtman grants Lewis that psychological similarity and interrelatedness of person stages keeps the *contents* of consciousness stable.

“The moments of consciousness themselves, however, remain distinct from one another, and what [Lewis] describes is more like a Humean bundle than a real unity. [...] My relation to my future self, on this view, is like my relation to someone very like me psychologically, a kind of super psychological twin. Just as my psychological likeness to a twin does not make it legitimate to hold me responsible for her actions, [...] psychological continuity and connectedness does not make it legitimate to hold me responsible for the actions of my past self if that is all that her being me amounts to.”⁶¹

Without real unity, psychological continuity falls short of providing reasons why the distinct moments of consciousness should extend concern or moral responsibility to each other.⁶²

The second supposed practical absurdity of Lockean positions comes from the relations of continuity and connectedness. Critics argue that by themselves, these relations are unimportant. For example, Swinburne thinks that Lockeans must maintain problematic claims about death and killing. For them,

“to say that existing people would be deprived of future experiences is just to say there would not be persons whose bodies are continuous with those of the living, or who had similar memories and character to the living, having experience. But then what’s so important about the persons, who are to have experiences, having bodies continuous with persons now living or having similar memories and character? Surely *in itself* nothing at all. If instead there are newly bred persons with healthier bodies, happier memories, and more amiable characters than those murdered, why should killing be wrong just because of the lack of bodily and

⁶¹Schechtman 2014, p. 36.

⁶²Schechtman (1996, pp. 55-60) thinks this is especially challenging for four-dimensionalist accounts in which a person’s temporal parts are distinct from each other. For more on this claim, see further below (2.5.1). Schechtman (2014, ch. 1) offers a reading of Locke which she thinks survives the quoted objection.

other continuity? *In itself* surely such continuity has no value.”⁶³

Swinburne concludes that continuities can at best be *evidence* for what does have such value: personal identity.

Madell formulates a similar worry about anticipation:

“[I]f I fear that I shall suffer pain what I fear is, not that the person who suffers this pain will have a certain set of memory impressions and a certain set of desires and intentions, but, quite simply, that he will be *me*. [...] If one is told at this juncture by the supporter of the psychological criterion that the reason I should be concerned is that to have such memories, and so on, just *is* to be me, I am again left with the feeling that no explanation has been offered. [...] [I]t is not at all obvious that I have *any* reason to be concerned about the fact that the person who will be in pain will have a certain set of memory impressions, and it is no clarification at all to be told that in this sort of context that is all that being me involves.”⁶⁴

Moreover:

“the only tolerable answer to the question ‘Why fear that future pain?’ is that it is, unanalyzably, mine.”⁶⁵

For both Swinburne and Madell, continuities are less important and valuable than personal identity. They think it follows that continuities cannot be what personal identity consists in.⁶⁶

Parfit describes one rationale for these reservations.⁶⁷ Simple positions like Swinburne’s regard personal identity as a *deep* further fact beyond continuities. They believe that this further fact is what gives us reasons for future-directed concern. Continuity relations might have some derivative practical importance once the further fact obtains. But

⁶³Swinburne 1974, p. 246.

⁶⁴Madell 1981, p. 110.

⁶⁵*Ibid.*, p. 112.

⁶⁶I disagree further below (2.5.4).

⁶⁷Parfit 1984, pp. 308-9.

if one accepts Swinburne’s picture, continuities are no surrogate for the deep further fact.

2.2 The Extreme And The Moderate Claim

Parfit distinguishes two broader claims one could endorse in light of the foregoing objections.

The Extreme Claim “[I]f the Reductionist View is true, we have no reason to be specially concerned about our own futures.”⁶⁸

This claim concerns reductionism, which is closely related if not equivalent to the complex view (5.1.2). I assume here that the extreme claim also applies to the complex view. The alternative is:

The Moderate Claim Continuity relations (for Parfit: relation R) give us a reason for special concern.⁶⁹

These claims allow for different readings. Whiting distinguishes the following variants.⁷⁰ First, special concern for one’s own future can be either irrational or merely not rationally required. Second, as an *absolute* claim, there is *no* reason to care about my future selves. As a *comparative* claim, there is no reason to care about our own futures *more* than about the futures of others, or no *special* reason for such concern.⁷¹ Whiting thus recommends that we think of the extreme and moderate claims as “a family of claims”⁷². Whenever I speak of ‘future-directed concern’ in the

⁶⁸[Ibid.](#), p. 307.

⁶⁹[Ibid.](#), p. 311.

⁷⁰Whiting 1986, p. 549.

⁷¹Hummel 2012, p. 20.

⁷²Whiting 1986, p. 549.

following, I intend to be neutral between a *comparative* and an *absolute* kind of concern.

Proponents of the claims are not always explicit about which version they endorse. As quoted earlier, Butler suggests an *absolute* extreme claim, given that it would be fallacious to be interested in *anything* that will befall us tomorrow. Immediately thereafter, Butler makes the *comparative* claim that

“if the self or person of today, and that of tomorrow, are not the same, but only like persons, the person of today is really no more interested in what will befall the person of tomorrow, than in what will befall any other person.”⁷³

Moreover, Butler speaks of a *fallacy* to care about the future if Locke’s position was true. This suggests that he thinks the theory makes it *irrational* to have future-directed concern. Parfit seems to understand the extreme claim in a similar way when he draws parallels⁷⁴ between future-directed concern within the complex view and *Future Tuesday Indifference*,⁷⁵ i.e. ordinary future-directed concern, except for being wholly indifferent about any pains or pleasures on future Tuesdays—one of Parfit’s paradigm cases of irrational concern.

The extreme claim and the moderate claim are not explicit about the exact source of this rational status. For example, Street distinguishes attitude-*dependent* from attitude-*independent* conceptions of normative reasons.⁷⁶ According to attitude-*independent* accounts, there are facts about how an agent has most normative reason to live that hold inde-

⁷³Butler 1736, p. 102.

⁷⁴Parfit 1986, pp. 832-3.

⁷⁵Parfit 1984, p. 124.

⁷⁶Street 2009, p. 274.

pendently of that agent's evaluative attitudes in combination with the non-normative facts. According to attitude-*dependent* accounts, no such facts exist; an agent's normative reasons derive solely from her evaluative attitudes and what is entailed by them together with the non-normative facts. I believe that the extreme claim and the moderate claim are compatible with either conception. The rationality status of concern against the backdrop of the complex view can be seen as arising from facts that are independent of the agents' attitudes. For example, we can understand the *absolute* version of the extreme claim such that regardless of the agent's actual values and desires, she would make a mistake in caring about the future if the complex view was true. But alternatively, we could also understand the extreme claim as making a point about our actual evaluative attitudes. They presuppose something which the complex view cannot provide. Given the insight that there are no deep further facts of personal identity, caring for our futures becomes incoherent, relative to our antecedent attitudes.

D. Shoemaker⁷⁷ describes a distinctive sense in which a theory of personal identity can be expected to give reasons for future-directed concern and other practices. The theory shall not only capture and explain facts that are *necessary* for these practices and their reasonableness. We expect more: a satisfying theory of personal identity articulates conditions that are *sufficient* for the reasonableness of future-directed concern and other attitudes. The conditions shall *make it sensible* that we have special concern for those future person stages that are related to us as

⁷⁷D. Shoemaker 2007, pp. 321-2.

specified by the theory.⁷⁸ The suggested claim is not just that identity matters. Instead, accounts of personal identity can be expected to give us an idea about how and why identity is what matters.⁷⁹ It is plausible to understand the extreme claim as a complaint about the complex view's failure to *make it sensible* why we care about our futures.

The extreme claim is formulated and discussed in the literature mostly with respect to concern for our *futures*. It is fruitful to think about variations in the *attitude* and the *temporal direction* figuring in the extreme claim. The attitude of *concern* looks inherently future-directed, but there are also past-directed instances or analogues, e.g., concern about whether it was *me* who did a wrongful action. Related attitudes such as regret, pride, and feelings of responsibility are paradigmatically past-directed. The extreme claim in this more general form, covering several kinds of practices and temporal orientations, might be much more difficult to defend. D. Shoemaker argues that practices like concern, moral responsibility ascriptions, compensation, etc. are too diverse and not unified enough for there to be *just one* grounding relation between them and personal identity.⁸⁰ The plausibility of the extreme claim might differ across the practices. Parfit's formulation of the extreme claim is restricted to future-directed concern. In the following, I will focus on this version of the claim.

Parfit acknowledges that the extreme claim is defensible if one believes in deep further facts of personal identity. Compared to them, continuity relations seem unimportant. For example, in Swinburne's picture

⁷⁸For a similar point, see S. Shoemaker 1984, p. 71.

⁷⁹As I explain below (4.5), D. Shoemaker himself *discards* this claim (2016, p. 323).

⁸⁰D. Shoemaker 2007, pp. 353-4.

of fission, even if continuity relations branch symmetrically, the further fact of personal identity obtains for at most one of the offshoots. The other offshoot “will not be a mere stranger”⁸¹, and thus should not be treated by the fissioner like everyone else. But the fissioner should regard this offshoot as a “mere instrument”⁸², and can rationally will the offshoot’s death if it threatens to interfere with the fissioner’s projects. Within Swinburne’s picture, continuity relations are not enough to motivate special concern that normally goes with the deep further fact.

Moreover, the extreme claim is defensible if one does *not* believe in deep further facts of personal identity. In that case, Parfit believes that the relation to the offshoots is as good as ordinary survival. This claim is neutral on whether special concern is warranted in ordinary survival.⁸³

The moderate claim is defensible, too. We care specially about people we love even without a deep further fact of personal identity between us and them. However, Parfit is unsure if this does the trick:

“Suppose I learn that someone I love will suffer great pain. I shall be greatly distressed by this news. I might be *more* distressed than I would be if I learnt that *I* shall soon suffer such pain. But this concern has a different quality. I do not *anticipate* the pain that will be felt by someone I love.”⁸⁴

He thus agrees that the moderate claim can be denied. In conclusion, he suspends judgement on whether the extreme or the moderate claim is correct: “I have not yet found an argument that refutes either.”⁸⁵

⁸¹Parfit 1984, p. 309.

⁸²*Ibid.*, p. 310.

⁸³*Ibid.*, pp. 310-11.

⁸⁴*Ibid.*, p. 312.

⁸⁵*Ibid.*

I close with one constructive suggestion. If we draw Whiting's distinction between absolute and comparative versions, a possibility arises for *combining* the extreme claim with the moderate claim: comparative extremism (the complex view does not give us reasons for special concern for our own futures) is compatible with absolute moderatism (the complex view does give us reasons to care for our futures). Maybe Parfit struggles to choose between the extreme and the moderate claim because his preferred, particular versions of the claims are not in opposition. For example, he seems to sympathize with idea that absence of a deep further fact motivates *comparative* extremism: the relation between me and my future selves is less deep than we thought, and thus special concern for my future selves is unmotivated. Moreover, fission shows that identity is not what matters in survival. If anything does matter in survival, it is relation R, a combination of psychological continuity and connectedness. This motivates *absolute* moderatism: continuity relations do give us a reason for future-directed concern.

2.3 A Reductio Against The Complex View

The extreme claim does not by itself undermine the complex view. But it figures in arguments against the complex view. As described in (1.2.2), Butler, Madell, and others defend the following argument:⁸⁶

⁸⁶I use an *comparative* phrasing by speaking of *special* concern, but a similar argument could be formulated with an *absolute* phrasing.

2.4. RESPONSE 1: DENY A COMPLEX/SIMPLE ASYMMETRY 39

- (P1) If the complex view is true, we have no reason for special concern for our own futures.
- (P2) We have reasons for *special* concern for our own futures.
-
- (C) The complex view is false.

The conclusion is the result of a *reductio*. If we assume the truth of the complex view, there would be no reason for special concern. But since we *do* have reasons for special concern, the complex view must be false. In the following, I argue that the *reductio* against the complex view can be resisted.

2.4 Response 1: Deny A Complex/Simple Asymmetry

The *reductio* is supposed to be a challenge specifically for the complex view. I now characterize three mutually compatible ways to deny this.

2.4.1 The Simple View And Concern

Simple theorists must claim that they can give reasons for special concern. Otherwise, the *reductio* would not be a competitive disadvantage of complex positions, but apply to simple ones as well. As Johansson notes, this would render the extreme claim less threatening:

“[o]ne could hardly fault a theory for being unable to provide something that no theory can provide.”⁸⁷

However, simple theorists say little about why their positions are superior in justifying concern. Instead, they typically take for granted that the

⁸⁷Johansson 2007, p. 642.

persistence of one's soul or ego is necessary and sufficient for concern.

As mentioned, Parfit thinks the extreme claim is defensible, given that the complex view denies deep further facts. Pre-reflectively, we are inclined to believe in these facts, and our ordinary concerns presuppose them.⁸⁸ But Parfit does not explain what it is about these deep further facts that would motivate or demand concern.⁸⁹

One problem is that the more featureless souls or egos are taken to be, the less it is clear what about them could matter. Whiting suspects:

“my soul—if I've got one—is no different from anyone else's except in so far as it has my experiences, memories, intentions and so on. Without these things, my soul is nothing to me.”⁹⁰

Similarly, Wolf writes:

“given that a Cartesian ego is independent of personality, memories, even physical and psychological continuity, surely that is not a very strong or sensible ground for caring about someone. There is at least as much reason to care about individuals that are connected to my present consciousness in other ways.”⁹¹

If a soul is merely the locus of experiences, memories, intentions and the like, then souls of different people do not seem to differ much. They are just neutral bearers of properties and relations of continuity and connectedness. This is particularly obvious for the Featureless Cartesian View, where facts about the ego can completely float free from psychological or bodily continuity. For Whiting and Wolf, sameness of ego or soul would matter only derivatively, if it could guarantee or sustain psychological

⁸⁸Parfit 1984, p. 281.

⁸⁹S. Shoemaker 1985, p. 452.

⁹⁰Whiting 1986, p. 547.

⁹¹Wolf 1986, p. 707.

continuity. But because it does not, the importance of sameness of ego or soul is questionable.

Locke makes a similar point about action. He imagines that a present person has the same soul that was in Nestor or Thersites. Unless this person is also *conscious* of Nestor or Thersites' actions, he cannot be concerned about their actions or attribute their actions to himself.

“For this would no more make him the same person with Nestor, than if some of the particles of matter that were once a part of Nestor were now a part of this man; the same immaterial substance, without the same consciousness, no more making the same person, by being united to any body, than the same particle of matter, without consciousness, united to any body, makes the same person.”⁹²

With regards to action, even if all these people share a soul, as long as they are not conscious of each others' actions, it is *as if* their souls would be distinct. By itself, having the same soul is insufficient to warrant concern for past actions.⁹³

These criticisms concern souls' or egos' *qualitative* features, or lack thereof. A different aspect which could make egos or souls important is that their presence gives a sense in which *one and the same* subject has the different experiences, memories, and intentions, which according to the foregoing objection can vary so wildly. Even then,

“we might well worry that numerical sameness of subject alone is not a very reasonable basis for concern.”⁹⁴

The same goes for the suggestion that the presence of egos or souls provides—although not necessarily one and the same subject of experien-

⁹²Locke 1690, Book II, xxvii. 14.

⁹³Note how Swinburne's worry that continuities by themselves are unimportant is just an instance of this argument, but turned *against* the Lockean position.

⁹⁴Kind 2004, p. 543.

ces—at least one and the same *entity* that exists at different times. This by itself does not avoid *comparative* extremism; it does not explain what it is about one’s soul that requires prioritized concern over the soul of someone else.⁹⁵ Neither is it obvious how the presence of souls would avoid *absolute* extremism; the question remains what it is about souls that requires concern at all. More needs to be said on why souls and not continuity relations are proper objects of concern, and why death should seem bad only if it means cessation of the former. In other words, these suggestions only push back the problem. The question remains why numerical identity between present and future souls motivates future-directed concern.

Another problem is that the deep further facts of personal identity are epistemically inaccessible, even to the subjects themselves. After all, in Swinburne’s picture of fission, both offshoots would be convinced that they are identical to the fissioner, while only one can be right. It is not a particularly attractive view that such entirely inaccessible facts are the rational basis for concern.⁹⁶

Although the reasoning of simple theorists about how their positions justify future-directed concern is mysterious, I make one charitable suggestion. The persistence of her soul or ego might provide the present self with a sense of ownership of future experiences. If a future experience is mine, we still have not been told much about what this involves and which justificatory role it plays with regards to concern. But we could tentatively grant that a sufficient condition for my concern is satisfied.

⁹⁵Johnston 2010, pp. 176-8.

⁹⁶Gasser 2013, p. 278.

At this point, in order to determine whether this point is a competitive advantage of simple positions, the focus should shift to their complex opponents (2.5).

2.4.2 Sources Of The Reasonableness of Concern

The first premise of the *reductio* is the extreme claim: if the complex view was true, we would have no reason for special concern. I see two presuppositions under which the premise could be true: first, the complex view, or a theory of personal identity more generally, is the *only* potential source for such reasons.⁹⁷ Second, whatever other reasons there seem to be for special concern, coming to realize the truth of the complex view *cancel them out*.

One suggestion to dispute the premise and its presuppositions is provided by D. Shoemaker. He proposes that for some attitudes, the question of whether they are justified is a non-starter. For example,

“[l]ove, rather notoriously, seems to operate independently of, even against, rational considerations. A mother’s love for her son is, we say, unconditional: her love obtains (or should obtain) no matter what. [...] And if rational grounding is off the table where the love of others is concerned, surely it is even farther off the table where love of self is concerned. [...] asking the question about rational justification in either case fails to take seriously the deeply natural commitment we have to the practices as they stand.”⁹⁸

One inspiration for this position comes from Strawson. He claims that even in view of determinism, our reactive attitudes which ground the practice of holding each other responsible are “part of the general frame-

⁹⁷Kind 2004, p. 550.

⁹⁸D. Shoemaker 2007, pp. 348-9.

work of human life, not something that can come up for review”⁹⁹. Demanding justification of such attitudes is misguided. D. Shoemaker’s suggestion is that concern for our futures is similarly natural and non-rational—as opposed to irrational. Such attitudes are not the right kind of states to be backed up by reasons.

D. Shoemaker ends up discarding this suggestion. Even if we grant that loving family members and maintaining reactive attitudes are non-rational, many concerns *are* proper objects of rational evaluation.¹⁰⁰ Sometimes people do seem to make genuine mistakes about concern, for example in *Future Tuesday Indifference*, or when caring too much about themselves. These judgements presuppose that concern can be evaluated in light of reasons.¹⁰¹ Moreover, in tension with the non-rationality thesis, we can reason ourselves out of certain ways of caring. Buddhists and Stoics even manage to reason themselves out of concern *as such* in view of their descriptive and normative beliefs. D. Shoemaker agrees that it might be difficult to say something principled about the rational grounding of concern. Still, he concludes that given these examples, it is not attractive to regard instances of concern or the practice as such as non-rational.

Wolf seems to defend a D. Shoemaker-style non-rationality thesis, at least about special concern for our own futures. She thinks that the question why we should have such concern is one instance of the more general

⁹⁹Strawson 1962, p. 14.

¹⁰⁰D. Shoemaker 2007, pp. 349-52.

¹⁰¹Note also how D. Shoemaker’s quoted remarks about parental love are slightly puzzling. Given that his case rests on the basis that parents *should* love their children, the attitude seems to be at least somewhat responsive to normative reasons. The question then arises how it can be a paradigm case of non-rational concern.

question why we should care about some people more than others. In all these cases, Wolf is convinced that *justifying* special attachments is misguided. Any facts we might mention, such as family relations, a common history, or particular features of the person we care about, are neither suitable nor intended as justifications.¹⁰² Instead, Wolf argues that we do not typically think that these concerns are in need of justification.

“[T]he answer to the question, Why care about one person rather than another? may be that indeed, there *is* no reason. But on the other hand, we do not want or expect reason to govern here.”¹⁰³

Moreover, Wolf argues that it is conceptually and biologically impossible not to care about one’s future. Forming intentions and making rational decisions inevitably involve future-directed concern for oneself.

“The answer to the question, Why care especially about oneself? then, may be that one cannot help but care about that, and that, in any case, it should not be necessary to justify one’s special self-concern.”¹⁰⁴

According to Wolf, a slightly different question *is* reason-responsive: Why care about *persons*? Plausible answers will refer to the complexity of lives which persons can lead and distinguish them from other beings. Wolf also makes a vivid case that caring about persons as opposed to *R*-related beings makes the world a better place: “life, or if one prefers, the world, is better that way.”¹⁰⁵ Otherwise, friendships, parent-child relations, etc. would suffer and impoverish. For example, if what we value in our friends is *R*-relatedness, it seems we should be averse to our friends developing and undergoing psychological changes. This would

¹⁰²See also Adams 1989, p. 455.

¹⁰³Wolf 1986, p. 719.

¹⁰⁴*Ibid.*, p. 720.

¹⁰⁵*Ibid.*, p. 713.

seem bizarre. Thus, justifications can be given for why we should care about persons rather than *R*-relatedness. However, Wolf's point is that these justifications invite

“contemplation of a variety of issues with respect to which the metaphysics of personal identity plays no obvious role.”¹⁰⁶

To the extent that concern for persons can be backed up by reasons, personal identity is unimportant to this rational grounding.

A related picture is defended by Johnston. In his view, concern is rationally justifiable, but the justification mentions personal identity only in a minimal way. In particular, Johnston is convinced that whether the complex or the simple view is true is irrelevant. Like Wolf, he identifies concern for our own futures as part of a wider pattern of self-referential concerns that are directed at close persons and relationships. These concerns seem so fundamental and natural that Johnston is happy to accept them as “basic pattern[s] of concern”¹⁰⁷ which are *prima facie* reasonable. Just like certain fundamental beliefs, they enjoy

“a defeasible presumption in favour of their reasonableness; they cannot all be thrown into doubt at once, for then criticism would have no place from where to start.”¹⁰⁸

Johnston's only requirement is that such basic concerns are coherent and robust against informed criticism.¹⁰⁹ Concern for friends and family is reasonable independently of the metaphysics of friendship and family, and the same goes for concerns about future selves:

¹⁰⁶Wolf 1986, p. 708.

¹⁰⁷Johnston 1992, p. 599.

¹⁰⁸*Ibid.*

¹⁰⁹Hummel 2012, p. 37.

“[j]ust as these concerns require only the ordinary fact that one has friends and acquaintances, so self-concern seems only to require the ordinary fact that one exists and will exist.”¹¹⁰

For the justification of concern, the *ordinary fact* of personal identity suffices. The justification is agnostic on whether the complex or the simple view gives the right account of personal identity.

Sider articulates his view on the relation between metaphysics and practical concerns with similar terminology. He introduces a distinction between *ordinary facts* and claims about the *ultimate reality* which underlies ordinary facts.¹¹¹ Metaphysics is concerned with the latter. Ordinary facts can mention facts and entities which, according to our metaphysics, are false or do not exist. For example, even mereological nihilists, if liberal enough, can accept the truth of statements about tables if understood as claims about ordinary facts. But strictly speaking, for the nihilist ultimate reality does not contain tables. No composite objects occur at the most fundamental level.

Sider believes that a number of ordinary facts are entirely neutral on their underlying, fundamental reality. For example, the claim that

“(F) Princess Diana’s entire life would have been better for her, had she not died when she did, than it in fact was”¹¹²

is neutral on whether presentism or eternalism is true. And although Sider does not explicitly say so, we might add that (F) is true or false irrespective of whether or not ultimate reality contains persons and/or irrespective of how persons persist. Metaphysicians might disagree whether

¹¹⁰Johnston 1992, p. 600.

¹¹¹Sider 2013, pp. 157-8.

¹¹²*Ibid.*, p. 162.

Diana ultimately is a human animal or an essentially mental entity. They might disagree on whether what we *ordinarily* take to be her parts *really* compose a whole, or whether they are just mereological simples arranged a certain way. But no such claims about ultimate reality are being made by ordinary fact (F).

In this picture, it seems possible that ordinary claims like (F) are justified solely by virtue of other ordinary facts, none of which makes claims on the nature of ultimate reality. For example, Sider suggests that (F) can be accepted on the grounds that an event is bad if

“a person’s entire life would have been better for her if the state of affairs had not occurred than if it had occurred.”¹¹³

These grounds are metaphysically austere. If we extend this picture, we can argue that the reasons for concern about a future experience do not involve claims about the ultimate reality of personal identity. In particular, the explanation is neutral on how whether the complex or the simple view is true. It is not the case that if the complex view failed to give reasons for concern, our concerns would be unjustified. The reasons could still come from other sources, e.g., ordinary facts that are neutral on ultimate reality.¹¹⁴

The foregoing authors agree that the relevance of personal identity for the justification of concern is minimal at best. Kind accepts this,¹¹⁵ but develops a more permissive position. We may mention theories personal

¹¹³Sider 2013, p. 162. This suggestion comes from Feldman 1991, p. 215.

¹¹⁴Presumably, responses along the lines of Johnston and Sider are particularly attractive for animalists. Olson (1997, p. 70) and DeGrazia (DeGrazia 2005, p. 63) agree that the mere fact that a future being is the same biological organism as me falls short of giving me reasons for concern. These authors agree that further considerations are necessary.

¹¹⁵Kind 2004, pp. 550-2.

identity when giving reasons for special concern. For example, if we accept that concern for our present experiences is justified, the theories are informative about how to project this concern across time.

“As I move through time, I am going to carry my present-self special concern with me. Knowing that, I want now to direct my concern to the future and, in doing so, I want to apply it to the future self that is *me*. To answer the question ‘Which self is me?’ I rightly look to a theory of personal identity.”¹¹⁶

Similarly, Johansson maintains that even if we accept that we are justified on caring about our own futures, and that theories of personal identity are irrelevant to this justification,

“[w]e still have to enter into the debate between different accounts of my persistence conditions [...] in order to find out *which* future individual I am justified in caring specially about”¹¹⁷.

Pace Wolf, Kind and Johansson maintain that theories of personal identity *can* be relevant for justifying future-directed concern. Such theories help us to understand what it would mean to extend justified, present concern for ourselves across time. However, in this respect there is, again, no asymmetry between the complex and the simple view. Both tell us which future selves would be *us*.

Moreover, Kind emphasizes that which weight this insight has depends on us. Whether it will be *me* who has a future experiences might be *relevant* to whether I am justified in caring about the experience.¹¹⁸ But acceptance of this link between personal identity and concern imposes no constraints on the *outcome* of our evaluation.¹¹⁹ In particular,

¹¹⁶[Ibid.](#), p. 547.

¹¹⁷Johansson 2007, pp. 654-5.

¹¹⁸Kind 2004, pp. 543-4.

¹¹⁹This resembles Korsgaard’s position that we *may* be guided by beliefs about

the link is compatible with us ending up *discarding* the suggestion that particular concerns or concern as a whole is justified. Our judgement depends on what we make of the fact that a future self will be *us*.

I conclude that the reasonableness of concern need not be based on the complex view, the simple view, or personal identity in general. It can have motivations that are entirely neutral on or external to these matters (for two examples, see 3.2). The truth of the complex view then would not diminish the reasonableness of concern.

2.4.3 Dependence Between Personal Identity And Concern

There is another problem with the premise that the complex view would render special concern unjustified. I just argued that there might be *other* justifiers that can take over. But let us grant that a theory of personal identity is indispensable for justifying special concern, and that the complex view indeed fails to give reasons (to be discussed in 2.5).

The *reductio* is a *modus tollens*: supposedly, the complex view implies that we have no reason for special concern; in fact, we do have such a reason; the complex view thus must be false. But note that alternatively, complex theorists could defend the following *modus ponens*: the complex view implies that we have no reason for special concern; the complex view is true; we thus lack a reason for special concern.

metaphysics, but that normative considerations determine which weight these beliefs carry: “the metaphysical facts do not obviously settle the question [of which motives to choose; P.H.]: I must still decide whether the consideration that some future person is ‘me’ has some special normative force for me. It is practical reason that requires me to construct an identity for myself; whether metaphysics is to guide me in this or not is an open question” (1989, p. 112). More on Korsgaard below (3.2.1).

Just as simple theorists say little on which aspects of their positions justify concern (2.4.1), it is difficult to reconstruct their grounds for preferring the *modus tollens* over the *modus ponens*.¹²⁰ They seem to regard our possession and the justifiedness of special concern as natural and basic, something which is taken for granted in everyday life and so ingrained in our practices that overthrowing it comes with serious costs. This perspective is shared by some complex theorists. For example, recall Johnston’s presumption that self-referential concerns are reasonable:

“One can fail to identify with [and thus fail to have non-derivative concern for; P.H.] one’s future self. But this will seem reasonable only if there is some considerable reason to inhibit the natural tendency to so identify, the natural tendency around which is built one’s concern that one’s own life continue, go well, and be worthwhile.”¹²¹

These authors would have serious reservations about assimilating special concern to irrational attitudes like *Future Tuesday Indifference*. The theory must render the concern sensible, or at least not remove justification. The conclusion of the *modus ponens*—we lack a reason for special concern—is not an option. In a slogan, the obviously justified practice constrains what can count as the right theory.

The truth of this slogan is not obvious if one considers alternatives. As we just saw, Kind defends a much weaker relation between concern and personal identity: by explaining what it takes to be identical to a future experiencer, theories of personal identity tell us how we can project concern about our present selves into the future. But in this picture, theories of personal identity can remain agnostic on whether the concern

¹²⁰Kind 2004, pp. 541-2.

¹²¹Johnston 1997, p. 157.

that gets projected is justified in the first place. That is, these theories by themselves do not settle whether we have reasons—in the present or in the future—for caring about ourselves, and/or caring about ourselves in a special way. The picture is compatible with the *modus ponens* described above. Although there is a link between personal identity and the justifiedness of concern, the slogan that practice constrains theory is too simple. Things can go the other way round. We might be conceding too much if we accept the *modus tollens* and debate whether the complex or the simple view justifies special concern better. It is at least an option to conclude, on the basis of endorsing the complex view, that special concern is unjustified. Even if the extreme claim is accepted, we need not reject the complex view.

Parfit seems to reject the slogan. His imaginary cases convince him that a revision of our theory of personal identity is warranted. “The truth is very different from what we are inclined to believe. Even if we are not aware of this, most of us are Non-Reductionists.”¹²² But the simple view is false, and the complex view is true. For Parfit, this has consequences for how we should think about our futures.

“Is the truth depressing? Some may find it so. But I find it liberating, and consoling. When I believed that my existence was a such a further fact, I seemed imprisoned in myself. My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air. There is still a difference between my life and the lives of other people. But the difference is less. Other people are closer. I am less concerned about the rest of my own life, and more concerned about the lives of others”¹²³.

¹²²Parfit 1984, p. 282.

¹²³*Ibid.*

With the extreme claim and the *reductio* against the complex view, Parfit shares the premise that there is an important connection between theories of personal identity and the reasonableness of concerns. But Parfit disagrees on which sphere has *priority* and ultimately restricts the other.¹²⁴ If the right theory of personal identity fails to ground special concern, then we should get rid of the concern. D. Shoemaker formulates the following guiding assumption:

“whatever turns out to be the correct account of personal identity will fix our practices and concerns accordingly.”¹²⁵

I am not defending either Parfit’s or Kind’s picture of the relation and priority between personal identity and concern. What matters for my purposes is that there are plausible alternatives to the particular supposition—the slogan—that is implicit in the *reductio* against the complex view. More needs to be said by proponents of the *reductio* to rule out a *modus ponens* with the conclusion that concern is unjustified.

2.5 Response 2: Continuity Relations Matter

One obvious way to resist the *reductio* against the complex view is to dispute the extreme claim that within the complex view, there are no reasons for special concern. Given the objections outlined earlier (2.1),

¹²⁴As I will explain later, once our theory is revised and Parfit’s position accepted, he thinks that the *modified* concerns do not track identity anymore. That is, for Parfit the priority of matters of personal identity is compatible with the claim that upon reflection, identity is not what matters.

¹²⁵D. Shoemaker 2007, p. 319. He is not explicitly attributing this assumption to Parfit.

I will pursue two lines for responding to the extreme claim. Firstly, one could deny the numerical distinctness charge launched by Butler and others. Secondly, one could accept the numerical distinctness charge, but maintain that the distinctness between stages of one and the same person is innocuous.

2.5.1 Against Numerical Distinctness: Persistence

Theories of persistence can deny the numerical distinctness objection in different ways. Three-dimensionalists believe that objects, including persons, are spread out in space, and are wholly present at each time of their existence. Four-dimensionalists maintain that objects are spread out in both space and time, and have temporal parts at each time of their existence.

“Let us say that something persists iff, somehow or other, it exists at various times; this is the neutral word. Something perdures iff it persists by having different temporal parts, or stages, at different times, though no one part of it is wholly present at more than one time; whereas it endures iff it persists by being wholly present at more than one time.”¹²⁶

According to one version of four-dimensionalism, objects persist by perduring. Sider¹²⁷ and Hawley¹²⁸ introduce stage theory, or *exdurance*, as an alternative to perdurance. Stage theory shares with perdurantism the view that objects have instantaneous temporal parts. But contrary to perdurantism, these instantaneous parts rather than their composites are the satisfiers of sortal predicates. A person stage at a time, not the aggre-

¹²⁶Lewis 1986, p. 202.

¹²⁷Sider 1996.

¹²⁸Hawley 2001, ch. 2.

gate person composed of a maximal set of person stages, is a full-fledged person and the referent of the respective person name.

How would these theories address the numerical distinctness charge that within the complex view, persons at different times are similar at best, but never identical? I begin with endurantism. It is notoriously difficult to say what it means to be wholly present.¹²⁹ What is clear is that according to endurantism, if a person endures, *all of her* is wholly present at more than one time. Endurantists pride themselves of being able to make sense of the claim that

“[a]n enduring object that exists at one time is identical to itself existing at another time.”¹³⁰

It is not the case that two similar but distinct things, e.g., distinct temporal parts, are present at these times. It is thus built into the definition of endurantism that objections along the lines of the numerical distinctness charge against the complex view do not apply.

The question arises whether endurantism is compatible with the complex view. Merricks denies that endurantism can be combined with psychological continuity theory.¹³¹ Endurantists understand personal identity as *numerical identity* between a person wholly present at one time and a person wholly present at another time. Merricks thinks psychological continuity relates *distinct* psychological states at different times.¹³²

¹²⁹For some discussion, see Merricks 1994, pp. 180-2, 1999b, pp. 423-4; Sider 2001b, sect. 3.3; Crisp and Smith 2005; Hofweber and Velleman 2011.

¹³⁰Merricks 1994, p. 166.

¹³¹Merricks 1999a.

¹³²I am not convinced by Merricks' case that the relata of psychological continuity cannot be *persons*. He seems to think that psychological continuity with myself precludes variation in my psychological states, “[f]or a person has all and only the psychological states that she has” (*ibid.*, p. 985). Could a person not be related *to herself* by psychological continuity if she has one set of states at one time, and a different

Numerical identity is neither the same as non-branching psychological continuity nor can be analyzed in terms of it. Because of this, psychological continuity either cannot be regarded as necessary and sufficient for personal identity, or can be regarded as necessary and sufficient without thereby offering a *reduction* of personal identity, i.e. the latter would remain a *further fact* beyond the proposed necessary and sufficient condition—something which psychological continuity theory supposedly seeks to avoid. A middle way would be that personal identity is *constituted* by non-branching continuity. But Merricks rejects this suggestion on at least three grounds.¹³³ First, the constitution relation is mysterious.¹³⁴ Second, without much argument he asserts that even if we concede that there are paradigm cases of constitution, e.g., statues and bronze, personal identity and psychological continuity are disanalogous to these cases. Third, numerical identity is primitive, not constituted by anything. Although Merricks targets psychological continuity theory, his points seem to translate to accounts of personal identity that focus on bodily continuity.

Merricks' discussion has received critical responses.¹³⁵ I am defending neither endurantism nor its compatibility with the complex view. For my purposes, it suffices to note that if complex theorists can respond to Merricks' criticism and endorse endurantism, they would circumvent the

but continuous one at another time? This matters because it undermines Merricks' claim that personal identity and psychological continuity must have different rela-

¹³³Merricks 1999a, pp. 992-3.

¹³⁴We could refer Merricks to the suggestions in (chapter 6).

¹³⁵Rea and Silver (2000) argue that if Merricks was right, a parallel argument would establish that *perdurantism* is incompatible with psychological continuity theory. Merricks disagrees (2000). Brueckner (2009) offers an endurantist formulation of psychological continuity which he thinks is perfectly defensible. Langford (2014) argues that both Merricks and Brueckner are mistaken.

numerical distinctness charge.

Let us now turn to four-dimensionalism. For the perdurantist, person stages at different times are distinct, but can be stages of one and the same person. This picture aligns neatly with the complex view. A person's temporal parts are typically seen as causally continuous and connected, and so can instantiate the relations which the view takes to be constitutive of personal identity.¹³⁶ For example, Lewis calls the relation which unites the stages of one single person the *I-relation*, and suggests that it can be analyzed in terms of mental continuity and connectedness.¹³⁷

The perdurantist has a response to the numerical distinctness charge. On the one hand, she can deny the charge for *persons* and other maximal aggregates of object stages. The person to whom the person stages belong is one and the same at each time of her existence. On the other hand, she must accept the numerical distinctness charge for person *stages*. The temporal parts of one and the same person are distinct. However, the perdurantist can go on to deny that this implies that one stage has no reason to care about other, I-related person stages, or has no reason to care about them more than about non-I-related stages. Both stages are temporal parts of one and the same *person*. The *stages* are distinct, but they need not be left without reasons to care about each other.¹³⁸

¹³⁶The picture is even compatible with the simple view. Persons and their souls or egos, and maybe even Madell's irreducible self, can have temporal parts.

¹³⁷Lewis 1976, p. 21.

¹³⁸For more discussion on the interests of person stages, see 3.1.2 and 3.1.3.

The locus of concern in Lewis' proposal is a person stage. Common-sense concerns are about persons, not stages.¹³⁹ This is one motivation for stage theory, or exdurantism. The stage view seems to be compatible with the complex view for the same reasons as perdurantism. It shares Lewis' stance that person stages can be I-related, but is distinguished by the claim that instantaneous stages are full-fledged persons. Although they are instantaneous, they satisfy tensed predicates by virtue of being I-related to stages at other times—just as in Lewis' counterpart theory, a world-bound individual is possibly *F* by virtue of having an other-worldly counterpart that is *F*.

Sider explicitly deals with a version of the numerical distinctness charge. He clarifies that the stage view does *not* have the consequence that “statements that look like they are about what once happened to me are *really* about what once happened to someone else.”¹⁴⁰ Although satisfaction of a tensed predicate is analyzed in terms of satisfaction of a non-tensed property, the possessor of the tensed predicate is the present person stage.¹⁴¹ If you face pain tomorrow, the person stage experiencing the pain is distinct from your present self. But this only *appears* to invite a numerical distinctness worry. Your present self satisfies the predicate of going-to-be-in-pain-tomorrow.¹⁴² She thus need not be concerned about the predicates that are satisfied by someone else; she will be concerned

¹³⁹Sider 1996, pp. 435-7. In this context, consider also Wolf's claim, aimed at Parfit, that momentary person stages cannot be the primary objects of practical concerns: “If the reason we care about persons is that persons are able to live interesting, admirable, and rewarding lives, we may answer that time slices of persons, much less experiences of times slices, are incapable of living lives at all” (1986, p. 709).

¹⁴⁰Sider 1996, p. 437.

¹⁴¹See also Hawley 2001, sect. 2.6.

¹⁴²See also Johansson 2007, p. 652.

about the predicates *she herself* satisfies.

I conclude that endurantism, perdurantism, and exdurantism all provide promising resources for responding to the numerical distinctness objection. These theories explain how persons persist such that *they*, and not a distinct entity, will have certain experiences in the future. The theories are neutral on how concern is justified, and so will not by themselves refute the extreme claim. But provided we are entitled to care about *our* experiences, these theories make sense of how *future* experiences will be *ours*—even when combined with the complex view.

2.5.2 Quasi-Anticipation

Suppose that the numerical distinctness charge goes through. Does this mean that caring about our futures is unjustified?

We do maintain reasonable concerns for persons distinct from us, e.g., our loved ones. This suggests that even if we grant that the complex view is committed to numerical distinctness of persons at different times, it is implausible to endorse the *absolute* extreme claim that we have no reason for *any* future-directed concern. At the very least, we seem entitled to extend the unproblematic concern we maintain for our loved ones to our own futures.¹⁴³

Still, there might be a shortcoming. When assimilating concern for our own futures to concern for others, we seem to ignore some distinctive features of concern for ourselves. As quoted earlier, Perry illustrates:

¹⁴³One might support the *absolute* extreme claim on the grounds that even caring for loved ones is unreasonable within the complex view, given that their present selves are numerically distinct from their future selves. Besides pointing at the aforementioned theories of persistence, I would refer the objector to (2.5.4).

“Most of us have a special and intense interest in what will happen to us. You learn that someone will be run over by a truck tomorrow; you are saddened, feel pity, and think reflectively about the frailty of life; one bit of information is added, that the someone is you, and a whole new set of emotions rise in your breast.”¹⁴⁴

One of Butler’s points was that the complex view renders this reaction unjustified; you should be “no more interested in what will befall the person of tomorrow, than in what will befall any other person.”¹⁴⁵

How is concern for our own futures special? Parfit writes:

“Suppose I learn that someone I love will soon suffer great pain. I shall be greatly distressed by this news. I might be more distressed than I would be if I learnt that I shall soon suffer such pain. But this concern has a different quality. I do not anticipate the pain that will be felt by someone I love.”¹⁴⁶

Johnston notes:

“The vivid sense that one will *oneself undergo* certain experiences in the future gives one’s future-directed self-concern its special and urgent quality”¹⁴⁷.

These authors think that concern for ourselves has a distinctive phenomenology. Parfit mentions *anticipation*, which D. Shoemaker characterizes as

“a matter of having certain beliefs or expectations about what it will feel like to undergo a certain set of expected experiences. It thus includes both doxastic and phenomenological components.”¹⁴⁸

Special concern for our own futures and its distinctive phenomenological quality seem to be related to anticipation. In the Parfit quote, lack of

¹⁴⁴Perry 1976, p. 67.

¹⁴⁵Butler 1736, p. 102.

¹⁴⁶Parfit 1984, p. 312.

¹⁴⁷Johnston 1992, p. 600.

¹⁴⁸D. Shoemaker 2007, p. 343.

anticipation is taken as indicative of a lack of special concern. Sometimes anticipation and special concern are used interchangeably.¹⁴⁹ Velleman directly ties the concern's importance to anticipation:

“To imagine a future pain, for example, as it will feel in the psychological wake of my hereby imagining it is to do more than just imagine it. It's to imagine the pain as befalling a mind that has somehow been prepared by this very prospect of its occurrence. And to imagine a pain as experienced by a mind hereby so prepared for it is already to brace for the pain, to shrink from it, or to be otherwise caught up in it in some way. Anticipation that's cognizant of its effect on the prefigured experiences is thus a form of mental engagement with them that, to some degree, already constitutes their mattering.”¹⁵⁰

In this picture, one condition for special concern with regards to a future experience is a particular kind of anticipation—one in which the experience is imagined as prefigured by the anticipation itself.¹⁵¹

We are considering whether in view of the numerical distinctness charge, the complex view could simply assimilate concern for our own futures to concern for others. The foregoing suggests that it cannot. Concern for our own futures has special features. Concern for others might be justified, but this does not show that the distinctive concern for our own futures is justified. The *comparative* version of the extreme claim remains a live option.

It is striking that Parfit and others merely *assert* that concern for our own futures is special. Perry claims that once I learn that *I* am the individual in the truck case, a whole new set of emotions rise in my

¹⁴⁹For example *ibid.*, p. 320.

¹⁵⁰Velleman 1996, p. 73.

¹⁵¹Velleman argues that what counts as *our own* histories and futures in this context is not tied to identity, but to reflexive, first-personal thought. “[S]ome of the deepest concerns expressed in terms of personal identity are actually perspectival concerns about the self” (*ibid.*, p. 67).

breast.¹⁵² This is doubtlessly true. But it is also true that once I learn that it is *you*, or a third person, who gets run over by a truck, a whole new set of emotions rise in my breast. The question whether empirical evidence supports the claim that concern for our own futures is special is surprisingly neglected.

Of course, it might be a conceptual truth about *anticipation* that we anticipate only our own experiences.¹⁵³ More substantive would be the claim that one can *quasi-anticipate* only one's own experiences, where *quasi-anticipation* is just like *anticipation* but does not conceptually require identity between anticipator and experiencer.¹⁵⁴ Do we quasi-anticipate only our own experiences?

We can certainly think of circumstances where quasi-anticipation could occur. Although denied by Parfit, concern for family members or loved ones looks like a promising example. Of course, we do not quasi-anticipate the experiences of just *anyone*. Quasi-anticipation might well require particular relations obtaining between the anticipator and the future self, or at least an *expectation* of the anticipator that such relations obtain. One candidate relation is psychological continuity. Indeed, Whiting points at two ideas put forward by psychological continuity the-

¹⁵²However, according to Perry, the importance of identity and the justification of special concern for our own futures are only *derivative* upon our future selves' ability to carry out our present projects. He accepts that apart from *ego projects* which essentially involve ourselves, e.g., that *I* raise my children, we would be perfectly justified in caring specially about others if we could be sure that they advance our projects just as well as our future selves. But this latter condition is satisfied only in outlandish cases; Perry 1976, pp. 80-4.

¹⁵³For example: "anticipation is a necessarily reflexive state, because one can only anticipate what is conceived to lie in ones own future" (Higginbotham 1995, p. 237).

¹⁵⁴This suggestion parallels S. Shoemaker's introduction of *quasi-memory* (1970, p. 271) when discussing whether the memory criterion for personal identity is circular.

orists.¹⁵⁵ First, there are scenarios in which the degree of psychological connectedness between two person stages is just below the threshold that is necessary for personal identity. Second, entitlement to quasi-projective imagination and anticipation is proportional to the psychological distance between the anticipator and the future self. As a result, it is sometimes plausible to care for others more than for one's own future selves.

“This sort of graduated concern, though difficult to achieve given current cultural practices, seems to me both psychologically possible and phenomenologically plausible: the difference between the sort of anxiety I might have about giving a talk (or having an operation) *tomorrow* and the sort of anxiety I might have about giving a talk (or having an operation) ten years from now is, I think, potentially similar to the difference between the sort of anxiety I might have about my daughter's impending performance (or operation) and the sort of anxiety I might have about my niece's impending performance (or operation).”¹⁵⁶

Note that Whiting not only claims that we sometimes quasi-anticipate. She also defends the idea that these quasi-anticipations are perfectly justified. Sometimes, what justifies anticipatory concern is present *across* lives more than *within* a life.

Similarly, Persson argues that a fissioner can quasi-anticipate the experiences of her offshoots regardless of whether or not she is identical to them. “It is, then, not identity that provides experiential anticipation with its distinctive character”¹⁵⁷.

Given these possibilities, it is surprising that Parfit- and Perry-style assertions that we only quasi-anticipate our own experiences are not backed up by empirical evidence. In fact, there are neuroscientific ex-

¹⁵⁵Whiting 1999, pp. 455-60.

¹⁵⁶*Ibid.*, pp. 459-60.

¹⁵⁷Persson 2005, p. 316.

periments, e.g., by Ehrsson¹⁵⁸ and others,¹⁵⁹ which examine the effects of centering of visual images to the perspective of the viewer. Through video goggles, individuals receive visual inputs whose perspectival presentation brings about feelings of identification with the experiencer in the imagery. In some experiments individuals receive visual images that show other bodies, but filmed from the perspective from which the individuals would see their own body. Observed reactions, self-reports, and even brain scans suggest that the individuals quasi-anticipate touches and pains to the bodies in their visual fields as occurring to themselves. Experiments like these might not demonstrate more than that through illusions, subjects can briefly be tricked into quasi-anticipating experiences of individuals that are in fact distinct from the anticipating subject. The doxastic attitudes involved in anticipation are fallible; anticipation can involve mistaken beliefs about which future experiences will be mine.¹⁶⁰ It remains an open question whether quasi-anticipation occurs. Still, the role of such empirical evidence could be to invite further justification of the claim that quasi-anticipation and special concern can only be directed at one's own future selves.

Velleman offers an account of how quasi-anticipation can fall short of grounding special concern, or “what matters about future selves”¹⁶¹. For example, when quasi-anticipating the experiences of fission products, we fail to have the sort of intimacy we have with real future selves. This is because anticipation would need to be tied to the perspective of a

¹⁵⁸Ehrsson et al. 2007.

¹⁵⁹For an overview on this stream of research, see Blanke 2012.

¹⁶⁰Recanati 2007, p. 158 makes the analogous point about quasi-perceptions.

¹⁶¹Velleman 1996, p. 76.

particular fission product, and therewith involve one thought too many:

“I could no longer think just about how the future would look; I’d have to think about how it would look to particular, specified observers. [...] [T]here would be no life that I could anticipate without first picking it out for the purpose of projecting myself into it. Surely, a position from which I must deliberately project myself into a life is not a position on the inside of that life.”¹⁶²

The idea seems to be that deliberate and conscious projection precludes the immediacy and urgency of anticipations, features which rest on the anticipator’s “ability to regard their subjects unselfconsciously as ‘me’”¹⁶³. Quasi-anticipation does not guarantee this ability.

Similarly, Recanati thinks although quasi-anticipation is coherent, emotions like delight, remorse, guilt, and pride are in fact caused by modes of memory and anticipation which “[guarantee] that the remembered actions are actions *we* performed”¹⁶⁴. Recanati thinks that if these attitudes were caused by quasi-memory and -anticipation, they would be much weaker than they actually are.

I have two reservations about these ways of arguing that quasi-anticipation is not what justifies special concern. First, it is not obvious to me that quasi-anticipation is less immediate than genuine anticipation. For example, it is hard to see why just because concern for one of my fission products presupposes that I first determine which fission product’s perspective I am occupying while anticipating, the urgency, quality, and importance of my concern will—or should—be mitigated. For one thing, even with ordinary future-directed concern, something similar is necessary: that I specify some time-frame, e.g., whether I am concerned

¹⁶²[Ibid.](#), pp. 75-6.

¹⁶³[Ibid.](#), p. 73.

¹⁶⁴Recanati 2007, p. 163.

about my 40-year old self or my 80-year old self. Moreover, after having picked out the perspective of my fission offshoot who will experience pain, while the other offshoot receives pleasure, it is just not obvious why my quasi-anticipation, -fear, and their urgency would be reduced by the prior process of selecting this particular offshoot's point of view. Recanati's response reveals how worries along these lines do not fully appreciate the suggestion of quasi-attitudes. Per hypothesis, quasi-attitudes and their genuine counterparts are phenomenologically *perfectly alike*. If one worries that quasi-anticipations are weaker and less urgent than genuine anticipation, one is not playing the game of entertaining the possibility of quasi-anticipations. This insight is not conclusive, but it suggests that the aforementioned responses change the topic.

Secondly, suppose we grant that genuine anticipation is more urgent and immediate than quasi-anticipation. The question remains whether we are *justified* in caring about ourselves with special immediacy and urgency. Indeed, Whiting argues that although we actually operate with anticipation rather than its quasi-counterpart,

“this may be an artifact of a deep-seated evolutionary or cultural bias toward something like the Simple View.”¹⁶⁵

It is one thing to argue that we anticipate rather than quasi-anticipate, but another thing to establish that we are justified in caring specially about ourselves.¹⁶⁶

I began by accepting the numerical distinctness charge against the complex view. But I have characterized ways how special concern still

¹⁶⁵Whiting 1999, p. 459.

¹⁶⁶See also Parfit 1984, p. 308, and Rachels and Alter 2005, pp. 314-5.

makes sense in this picture. Anticipation is what grounds special concern, and quasi-anticipation can obtain across distinct persons. The numerical distinctness charge against the complex view does not establish the extreme claim that if the complex view was true, we would not be justified in caring about future experiences, and caring about some future experiences more than others.

2.5.3 Insights From Fission

For Parfit, fission suggests that identity is not what matters. Here, relation R branches symmetrically. He discusses a number of different descriptions of the case, e.g., that the fissioner does not survive, or that she survives as one or even both of the offshoots. Each of them has shortcomings, but Parfit thinks that fission is most plausibly described as a case of non-identity.¹⁶⁷ But this, Parfit argues, actually addresses an empty question. The options are not distinct possibilities, but different descriptions of one and the same possibility. Even without privileging one description, all the facts are clear. The much more interesting question is whether fission secures what matters. Parfit thinks it does.¹⁶⁸ What matters would be secured if there was only one offshoot.

“The relation between me now and that future person is just an instance of the relation between me now and myself tomorrow. So that relation would contain what matters. In the Double Case [with two offshoots; P.H.], my relation to that person would be just the same. So this relation must still contain what matters. Nothing is missing. But that person cannot here be claimed to be me. So identity cannot be what matters.”¹⁶⁹

¹⁶⁷Parfit 1984, pp. 255-60.

¹⁶⁸*Ibid.*, pp. 261-3.

¹⁶⁹Parfit 1995, p. 42.

Parfit shares this view with S. Shoemaker, who argues that non-identity between fissioner and fission products does not make the fissioner's future-directed concern and the fission products' backward-looking attitudes of remorse or pride less reasonable.¹⁷⁰ Olson thus speaks of the "*Parfit-Shoemaker Thesis*"¹⁷¹ about personal identity and what matters.

One can deny that the symmetrical branching of continuity relations is best described as a case of non-identity. As mentioned, some simple theorists believe that when continuities branch, the person does not really split. Instead, her soul or ego goes with one of the offshoots (2.2, 6.3). We also just saw Merricks' denial that personal identity can be analyzed in terms of psychological continuity. If so, Parfit is not entitled to regard non-branching psychological continuity as personal identity, and hence to equivocate the claim that non-branching psychological continuity is unimportant with the claim that personal identity is unimportant.¹⁷² A different objection comes from Parfit's fellow complex theorists. Lewis holds that prior to branching, there were as many persons as there are now branches. These persons shared all their spatial parts.¹⁷³ The branching thus does not bring about non-identity; it merely separates entities which so far overlapped spatially.

This is a point of agreement between at least some paradigm cases of complexity and simplicity. Since in Swinburne's and Lewis' pictures, identity and continuities never diverge, fission does not provide evidence for the irrelevance of identity even if one grants that it secures all that

¹⁷⁰S. Shoemaker 1970, p. 284; 1984, pp. 130-2.

¹⁷¹Olson 1997, p. 54.

¹⁷²Merricks 1999a, pp. 994-7.

¹⁷³Lewis 1976, pp. 24-9.

matters. In a sense, these theorists share the view that fission cannot occur. Whenever it seems like one person divides into two or more, this seeming can be explained away.

These objections to Parfit rest on particular *descriptive* pictures of what happens in fission. Another class of objections accepts Parfit's description, but rejects his *evaluative* judgement that fission is as good as ordinary survival. As a first example, Wolf points out that the fission offshoots would have equal claims for the resources of a single life. Awkward questions arise about how to balance these claims. Should we divide the fissioner's resources amongst the offshoots? Should we declare one of the offshoots the fissioner's successor, and ask the other to begin a new life? Either way, at least one offshoot would be worse off.

“[O]ne has strong reason to prefer *nonbranching* psychological continuity and connectedness. That is, one has strong reason to prefer *personal identity*.”¹⁷⁴

As mentioned earlier, Wolf also speculates that the world would be worse if people would care about relation *R* rather than identity.

Second, Sosa denies that fissioner and each offshoot standing in the same *intrinsic* relations as in ordinary survival establishes that fission is as good as ordinary survival. *Extrinsic* features can matter, too.¹⁷⁵ The value of exclusive ownership, marriage, competitions, and prizes arises partly from the extrinsic fact that *no one else* instantiates the relation. Similarly, it is defensible to regard it significant that a second offshoot is present in fission.

¹⁷⁴Wolf 1986, p. 716.

¹⁷⁵Sosa 1990, pp. 319-20.

Finally, Johnston argues that if we concede that fission is as good as ordinary survival, it does not follow that relation R is what ordinarily matters. Circumstances in fission are extraordinary. Two presuppositions of ordinary concern are violated: that it is determinate who will be who, and that one has at most one R -continuous successor. When these presuppositions are violated, it might be defensible to care about relation R .

“But this is not because identity is never what matters. Instead, this is because caring in this way represents a reasonable extension of self-concern in a bizarre case.”¹⁷⁶

In other words, the outlandish case of fission is not informative about what matters ordinarily, only about reasonable local modifications of concern when ordinary presuppositions are not satisfied.

Even if we grant Parfit that these objections can be met, insights from fission do not suffice to refute the extreme claim. Fission might be as good as ordinary survival. But how good is ordinary survival?

2.5.4 Ordinary Survival In The Complex View

According to the objections described earlier (2.1), the value of ordinary survival would be diminished if the complex view was true. If the extreme claim shall be resisted, arguments making a positive case for the value of continuity relations are needed.

Parfit’s positive suggestion is that relation R , i.e. psychological continuity and/or connectedness,¹⁷⁷ is what matters:

¹⁷⁶Johnston 1992, p. 611.

¹⁷⁷Belzer (1996) notes that Parfit is not consistent about whether relation R encompasses psychological continuity, connectedness, or both. This introduces compli-

“What we value, in ourselves and others, is not the continued existence of the same particular brains and bodies. What we value are the various relations between ourselves and others, whom and what we love, our ambitions, achievements, commitments, emotions, memories, and several other psychological features. [...] I believe that, if there will later be some person who will be *R*-related to me as I am now, it matters very little whether this person has my present brain and body. I believe that what fundamentally matters is Relation *R*, even if it does not have its normal cause.”¹⁷⁸

These are suggestions about what we actually value. What we want to know is whether we are we *justified* in valuing relation *R*. As mentioned, Parfit thinks that relation *R* is metaphysically less deep than personal identity as understood by the simple view. Insofar as our concerns rest on the assumption of metaphysically deep facts, they are unjustified.

However, it is not clear at all how a lack of metaphysical depth should deflate future-directed concern. To begin with, the notion of metaphysical depth itself is unclear. Presumably, charitable rephrasals include that metaphysically deep facts are not grounded in other facts, or grounded in fewer facts than metaphysically less deep facts (6.1), and/or fundamental facts as opposed to derivative ones (6.2).

Once made precise, it is not clear why it should be rationally impermissible to value facts that are metaphysically less deep in just the same way as metaphysically deep facts can be valued. For example, Brink argues that future-directed concern can be understood as being

cations. In order to have a fission case, it seems enough that *continuity* branches; fissioner and offshoots need not be directly *connected*. If we accept that fission is as good as ordinary survival, Parfit is entitled to conclude that identity’s non-branching requirement is not significant relative to what matters. But fission by itself is silent on what exactly does matter, including whether it is continuity, connectedness, or both. Only connectedness is a matter of degree; continuity is not. Fission alone thus does not fully support Parfit’s further claim that what matters comes in degrees.

¹⁷⁸Parfit 1984, pp. 284-5.

attached to “*functional role*, rather than to its metaphysical or compositional *analysans*.”¹⁷⁹ Materialists and dualists alike should avoid pain, regardless of how they analyze pain states. Brink admits that sometimes, a metaphysical analysis does influence an entity’s role in our conceptual network. For example, learning that the gustatory properties of chocolate ice supervene on fat content may affect its desirability. However, the analysis provided by the complex view should not have such influence on future-directed concern. As Whiting puts it:

“It has always mystified me how reasons for concern are lost in the move from a pain’s being unanalyzably mine to its being analyzably mine.”¹⁸⁰

With Sosa¹⁸¹ and Johnston¹⁸², we can add that this holds even if we accept that facts and relations mentioned in the analysandum appear less important than the analysans.¹⁸³ Neither the simple theorists nor Parfit have said enough about how the insight that personal identity consists in continuity relations makes caring about our futures less reasonable.

Note that these points need not be taken to show that identity is not what matters. All that has been suggested is that the view that personal identity is constituted by non-branching continuity relations does not diminish the importance of survival.

This suggests that *absolute* extremism—the idea that the complex view leaves us with no reason whatsoever for caring about the future—is

¹⁷⁹Brink 1997a, p. 117.

¹⁸⁰Whiting 1986, p. 552.

¹⁸¹Sosa 1990, pp. 320-1.

¹⁸²Johnston 1997, p. 167.

¹⁸³On this issue, see also Parfit’s *argument from below* (1984, p. 286, 1995, pp. 29-32, 2007, pp. 32-6) versus Johnston’s *argument from above* (Johnston 1997, pp. 167-8, 2010, pp. 307-16), and D. Shoemaker’s commentary (2006) on the exchange.

unmotivated. The complex view does not take away from our reasons for concern. Moreover, we could even deny *comparative* extremism and maintain that we have good reasons for special concern for our own futures. On the one hand, continuity relations give us no reason to care specially about our own future selves, where ‘our own’ signalizes identity. But on the other hand, we are entitled to care about some future selves—e.g., those *R*-related to us—more than others. This concern does not require identity. But it is *special* by virtue of being based on *R*-relatedness between the caring, present self and the future self that is the object of concern. Indeed, consider again Chisholm and Swinburne’s position on fission. Suppose the fissioner knows in advance that the offshoot who inherits her soul receives pleasure, while the other offshoot gets tortured. It would be bizarre to deny that the fissioner lacks a reason to have special concern for the tortured future self.¹⁸⁴

These moves remain primarily *defensive*: the fact that personal identity consists in relations of continuity and connectedness *does not take away* from the value of ordinary survival and the justifiedness of concern. In the next chapter, I will examine detailed accounts of *positive* pictures about how continuity and connectedness *confer* value upon future selves.

2.6 Conclusion

The complex view supposedly has implausible practical consequences (2.1). First, according to the numerical distinctness charge, the complex view fails to offer an account of personal identity. It only captures

¹⁸⁴Johansson 2007, p. 654.

similarity amongst distinct person stages. Second, the relations which the complex view takes to obtain amongst person stages do not matter in the way identity matters. In light of these objections, one can endorse the extreme claim, and hold that the complex view gives us no reason for future-directed concern (2.2). Alternatively, one can respond by endorsing the moderate claim, and argue that the complex view does give us reasons for concern. I noted that these claims are ambiguous. In particular, authors are not always careful to distinguish *absolute* from *relative* instances of the claims, i.e. whether they pertain to any kind of concern, or some kind of special concern we have—or should have—for ourselves.

I then sketched a *reductio* argument advanced by some opponents of the complex view: if the complex view was true, we would be left with no reasons for concern; we do have such reasons; hence the complex view must be false (2.3). In the remainder of the chapter, I distinguished two broad ways to respond to the *reductio*.

First, by showing how it is equally problematic for the simple view. If the simple view was true, we would not be left with reasons for special concern either. Alternatively, reasons for concern do not come from the complex or simple view anyways, so it is not a competitive disadvantage if one of them fares worse in justifying concern. Finally, it is defensible to maintain that our concerns *are* unmotivated, for example *because* the complex view is true and fails to give reasons for them.

Second, by denying the extreme claim. Theories of persistence can be combined with the complex view and give a sense in which the numerical distinctness objection—one of the motivations of the extreme claim—is false. And even if numerical distinctness cannot be eliminated

from the picture, the suggestion that we are sometimes justified in quasi-anticipating the experiences of others shows how this need not deprive us of reasons for concern. Although fission is one example for such cases, I argued that even if we read Parfit charitably and grant that objections to both his descriptive and evaluative remarks about fission can be addressed, it has only been shown that securing continuity relations is as good as ordinary survival. Finally, I argued that the value of ordinary survival is not diminished by the complex view.

As a result, we can reject the *reductio*. Concern for our futures is neither rendered unreasonable by the complex view nor constrains our theory choice. Let me note two further implications.

The discussion motivates more general challenges for positions according to which an everyday practice constrains what can count as an eligible theory. From proponents of such positions, we would want a precise account of the practice in question, what counts as a reason relative to the practice, and on which descriptive assumptions it rests. Moreover, we would need clarification on how these assumptions are in line with some theories more than others, and how certain theories cannot be made consistent with these assumptions. Only then can we appreciate a genuine conflict between engaging in the practice and endorsing a theory. For example, one lesson from Sider's discussion is that the assumptions of the practice might be coarse-grained and austere enough that none of the particular theories is in tension with them. Finally, we need independent motivations for the picture that practice is prior to theory and constrains what we can reasonably believe.

Moreover, Parfit's remarks on how the complex view lessened his fear

of death appear permissible. His antecedent reasons for caring about the future in specific ways are lost upon endorsing the complex view. But it is far from clear that everyone should rationally react in this way. The complex view preserves a robust sense of ownership of future experiences. And it is not clear why being related to future experiencers through continuity and connectedness rather than a metaphysically deep further fact—whatever this means—should reduce my immediate and urgent interest in their well-being. Leaving the glass tunnel might be liberating. But the complex view entitles us to such liberation only relative to a particular—and peculiar—conception of why we care about or futures.

Chapter 3

Grounds For Concern

In the previous chapter, I argued that the extreme claim is not a reason to reject the complex view. For the sake of the argument, I granted that within the complex view, present and future selves are distinct and cannot be connected by a deep further fact. It turned out to be unclear why this affects the value of ordinary survival. I now turn to the positive counterpart of this claim: what *does* matter in ordinary survival might be related to personal identity; but it is something else. We engage in forward-looking projects and attitudes not because someone will be *us*, but because we relate to future selves in other, more important ways.

The views I will consider endorse two broad strategies respectively. First, they make the complex view more precise. Personal identity consists in particular relations amongst person stages, and these relations are proper objects of concern. Second, they implement a suggestion described earlier (2.4.2) and look beyond the metaphysics of personal identity in order to give reasons for concern. Both strategies agree that what matters is not determined independently of us. We are involved in generating and constructing our relations to future selves. I will argue that the exact meaning of this claim needs to be clarified and qualified.

3.1 The Importance Of The More Particular Facts

Parfit characterized the complex view as taking personal identity to consist in more particular facts. Whiting and Brink are convinced that the relevant more particular facts encompass practically important relations

between person stages. Something that matters is constitutive of personal identity. This directly defeats the extreme claim. I first describe their positions, before offering reflections.

3.1.1 Whiting On Friends And Future Selves

According to the extreme claim, the complex view cannot give reasons for future-directed concern. Within the complex view, person stages are numerically distinct. Supposedly, this makes it mysterious why one should be concerned about the other (1.2.2). Moreover, Madell and Swinburne argue that psychological continuity by itself is unimportant (2.1). Whiting's aim is to address these objections on behalf of psychological continuity theory.

There are ways for complex positions to deny that person stages at different times must be distinct (2.5.1). Moreover, even if the complex view keeps person *stages* are distinct, this need not matter. The distinct stages can be part of one and the same *person*. But Whiting rejects appeals to numerical identity for justifying future-directed concern on the following grounds.¹⁸⁵ Suppose I face fission, and my fission offshoots will experience pain. If identity would be what provides reason for special concern, I have reason to care specially about the pain if only one offshoot survives. If two offshoots survive, I will not be identical to them, given that identity cannot be one-many. If identity is what matters, special concern would be unjustified. But it is quite implausible that the presence of a competitor affects whether I have reason to care specially.¹⁸⁶ I am

¹⁸⁵Whiting 1986, pp. 555-6.

¹⁸⁶This is an instance of the *Parfit-Shoemaker Thesis* (2.5.3).

entitled to such concern in both scenarios, yet only one is an instance of identity. Identity cannot be what grounds special concern. Thus, we need not even bother to reject the numerical distinctness charge in order to justify concern for our future selves.

Whiting's positive proposal is motivated by the following observation. Special concern extends not only to our future selves, but also to friends and family members. For example, special concern for our friends allows for interpersonal compensation: to some extent, benefits to our friends compensate *us*. Moreover, we seem *justified* in having special concern for our friends. If friendship justifies concern in the absence of identity, maybe psychological continuity with future selves can achieve something similar.

Whiting claims that what justifies concern for our friends is that we share experiences, desires, interests, values, and have interactions which shape these desires, interests, and values.¹⁸⁷ She notes that the same things give rise to psychological continuity between my present and future selves. For example, we are connected by memories, and share character traits, intentions, desires, and values. Friendship is a two-way relationship in which friends share, shape, and reciprocate these attitudes. On reflection, the relation to our future selves is similar. We obviously shape our future selves, e.g., through choices in the present. But our future selves also shape our present selves: we anticipate their desires and interests, and balance them with our current interests, e.g., when making planning efforts. Given these parallels, Whiting suggests that psychological continuity is to concern for our future selves just like

¹⁸⁷Whiting 1986, pp. 558-560.

friendship is to concern for our friends.

If psychological continuity was limited to memories, it would be passive and oriented towards the past. It would fall short of underpinning the persistence of *agents*. But psychological continuity encompasses more than just memories. In particular, Whiting focuses on *personal* desires. A desire is personal if it

“*essentially* includes its subject’s being in certain states or doing certain things”¹⁸⁸,

for example my desire that *I* raise my children. Two points about personal desires are central to Whiting’s position.

First, she suggests that psychological continuity theory can understand personal desires in a distinctive way: the desirer and the individual enjoying the desired state need not be numerically identical. All we need is that

“I desire the subjects under descriptions making use of first person indexicals [...] So I can have personal desires for the future satisfactions of subjects with whom I am not numerically identical.”¹⁸⁹

For the psychological continuity theorist, being *me*—and thereby being eligible to figure in my personal desires—means being psychologically continuous with me.

Secondly, such personal desires

“are primitive forms of concern for those subjects—that is, primitive forms of concern for my future selves.”¹⁹⁰

¹⁸⁸[Ibid.](#), pp. 561-2.

¹⁸⁹[Ibid.](#), p. 565.

¹⁹⁰[Ibid.](#)

With *primitive* concerns, Whiting means that *general concern* for our future selves derives from them¹⁹¹. Again, the analogy is friendship, which begins by taking particular interests in others. Whiting's example is an initial desire that my squash training partner wins her match. As time goes on, I learn more about her interests, projects, etc., and start to care about these things, too. As this progresses and she becomes my friend, *general concern* for her results: a second-order desire that her first-order desires are satisfied. According to Whiting's analogy, *general concern* for future selves arise in the same way from *primitive* first-order concerns, specifically personal desires that my future selves do and experience particular things.

But now note the following. Psychological continuity, which according to psychological continuity theory makes certain future selves ours, has concern for our future selves as a component. It is not the case that we are related to our future selves in a certain way, in this case psychological continuity, and derive reasons for concern from this relation. It is the other way round: we have primitive concerns qua personal desires which *give rise to* psychological continuity rather than deriving from it—just like friendship does not obtain independently of concern for friends, and reasons for concern for friends derive from it. Having, maintaining, and developing the concerns *constitutes* friendship.

“In this sense, we make our future selves by coming to care about them in much the same way that we make friends by coming to care about others.”¹⁹²

The fact that I am specially concerned about a future subject is part of

¹⁹¹Whiting 1986, p. 565.

¹⁹²*Ibid.*, p. 566.

what makes it my future self in the first place.

Whiting distinguishes two grades of psychological continuity and connectedness. A *thin* version obtains through passive states like memories and impersonal beliefs and intentions. *Personal* desires and the resulting intention-connections are components of *thick* psychological continuity. This distinction allows Whiting to express the following.

First, it is not a disanalogy between friends and future selves that the former exist independently of our concern. *Thin* psychological continuity gives a sense in which

“there will in the future be subjects who are psychologically continuous with us in ways which are relatively independent of our concern for them.”¹⁹³

This thin relation holds independently of whether I, through my special concern, make these subjects *my* future selves.

Second, Whiting avoids voluntarism about future selves.¹⁹⁴ My future selves are not simply the selves about whom I care specially. Just like my desire to be friends with you does not suffice for establishing friendship, my special concern about future selves does not suffice for making them *my* future selves. In both cases, further conditions need to be met. Only if my concern brings about intention-connections and other varieties of *thick* psychological continuity between my present self and the future self will the latter be mine. Since there are no *thick* connections between your future selves and me, my concern about your future selves does not make them *my* future selves.

¹⁹³[Ibid.](#)

¹⁹⁴Whiting 1999, pp. 465-6.

Whiting's account faces a circularity worry:¹⁹⁵ ordinarily, learning that a future subject will be *me* gives me reason for concern. But on the proposed account, this would be circular; the subject is my future self in virtue of me having the concern. It now looks as if I have reason to be concerned for my future self because I care about it in the first place.

Whiting responds that for friendship, the circularity doesn't seem so bad. It is fine to say that friendship gives us reasons for caring even if caring partly constitutes friendship. Moreover, the force of the circularity worry depends on whether we can give reasons for the initial or primitive concerns which eventually give rise to general concern for friends or future selves respectively. If we cannot give independent reasons for these initial concerns, then indeed all we can say about why we should have the general concern is that we are already caring in the first place. However, if the initial concern is itself responsive to independent reasons, then circularity could be avoided. Whiting suggests that for both friends and future selves, initial concern rests on our appreciation of character, projects, desires, and commitments of which we approve independently of our relationship to the friend or future self respectively. And once the initial concern thus motivated has constituted general concern, we get additional reasons to care about them.

If we follow Whiting, we can conclude that there is something fundamentally wrong with the objections (2.1) that motivate the extreme claim. It is not the case that because person stages at different times are distinct, they cannot be related by concern. And it is not the case that psychological continuity by itself is unimportant, given that personal

¹⁹⁵Whiting 1986, p. 570.

desires are components of psychological continuity.

If convincing, this shows that we should care about our future selves. But one remaining question is whether we should care for our future selves in a *special* way. After all, justification of concern for future selves is analogous to justification of concern for friends. This seems to allow that we should care about friends and future selves in similar ways, and not necessarily privilege *our own* futures.

Although her primary concern is to disarm the extreme claim as a reason to reject the complex view, Whiting actually agrees with one version of the extreme claim: we are not *rationally required* to care more about our own futures than others. She gives two reasons.

First, although maintaining at least minimal personal desires might be practically necessary, e.g., in order to form and act upon intentions,¹⁹⁶ we are not rationally required to have personal desires, and are free to renounce them entirely—just as becoming invested into characters and projects of potential friends is admissible but not required.¹⁹⁷

Second, if we maintain personal desires, there is no rational requirement to prioritize them over concern for others. Personal desires might be weak compared to at least some impersonal desires we have with regards to others,¹⁹⁸ e.g., when parents care more about their children than themselves. In that case, too, a comparative version of the extreme claim is appropriate, and there is a sense in which we need not care for ourselves in a special way.

¹⁹⁶*Ibid.*, pp. 568-9. As mentioned in 139, Wolf (1986, p. 719) is skeptical about the conceptual possibility of agents giving up all personal desires. Kind (2004, p. 548) argues that the personal desires one maintains can be extremely short-lived.

¹⁹⁷Whiting 1986, p. 575.

¹⁹⁸*Ibid.*, p. 576.

One vague worry about Whiting's picture is that it presupposes too much concern for others. If when seeking reasons for concern about my own future, I need to take for granted that I am justified in caring about others, some would wonder whether I have assumed something way more demanding than I set out to justify. In view of reservations along these lines, it is fruitful to turn to Brink's position. He justifies concern for our own future and concern for others in ways that resonate with Whiting's approach, but does so from the perspective of *rational egoism*.

3.1.2 Brink On The Separateness Of Persons

Brink's position on personal identity and special concern focuses on one distinctive way in which we can be concerned about our own futures: by having *prudential* concern. According to *prudence*, one should promote one's own overall interests through equal concern for all parts of one's life. *Rational egoism* is the view that all there is to practical reason is prudence.

Prudence faces a parity challenge:¹⁹⁹ it requires an agent bias but no temporal bias, i.e. the agent shall care specially about her own interests, but care equally about all times of her existence. The question arises why one should not distribute goods and harms by the same principles within lives and across lives. Parity considerations motivate the view that each of us should either favour herself now (*presentism*) or renounce any agent- or time-biases (*neutralism*).

In order to justify the hybrid structure of prudence, Brink refers to

¹⁹⁹Brink 1997a, pp. 98-102; Brink 2017, sect. 2. The parity worry goes back to Sidgwick 1907, p. 418, and is also stated in Nagel (1979, pp. 99-100) and Parfit (1984, p. 144).

3.1. THE IMPORTANCE OF THE MORE PARTICULAR FACTS 87

two thoughts of Sidgwick.²⁰⁰ First, in order to make sense of temporal neutrality, *prudence* must tell us why we should make sacrifices in the present if the benefits are enjoyed in the future.²⁰¹ The underlying thought is that if an individual makes a sacrifice, some form of appropriate compensation is needed to make the sacrifice rational. Secondly, the key to this is the separateness²⁰² of persons, i.e. the fact that “the distinction between any one individual and any other is real and fundamental”²⁰³. The separateness of persons justifies present sacrifices for a later benefit:

“In the intrapersonal case, benefactor and beneficiary are the same person, so compensation is automatic. In the interpersonal case, benefactor and beneficiary are different people; unless the beneficiary reciprocates in some way, the benefactors sacrifice will not be compensated. Whereas intrapersonal compensation is automatic, interpersonal compensation is not. This fact about compensation appears to rationalize intrapersonal neutrality without rationalizing interpersonal neutrality.”²⁰⁴

Some sense of personal separateness is presupposed in making sacrifices for the sake of future benefits. I am justified in having a bias towards myself in caring and planning because if the same separate person will enjoy the outcomes, compensation is *automatic*.²⁰⁵ What matters is that *I* receive the benefit; all things being equal, the temporal location of the benefit does not matter. This is how we arrive at *prudence*, the combination of agent bias with temporal neutrality.

²⁰⁰Brink 1997a, p. 105; Brink 2017, sect. 3.

²⁰¹Sidgwick 1907, p. 418.

²⁰²This notion of separateness is not to be confused with Parfit’s characterization of the simple view according to which persons are *separately* existing entities beyond physical or psychological states and relations (5.1.1).

²⁰³Sidgwick 1907, p. 498.

²⁰⁴Brink 2017, sect. 3.

²⁰⁵Brink 1997a, pp. 106-7.

Brink discusses objections. For example, one might be concerned that me-now and me-later are distinct, too. Why does this separateness not threaten intrapersonal balancing? His response is that there are several reasons to refrain from regarding person stages as agents.²⁰⁶ If we take momentary time slices of a person as the relevant units, the problem is that they do not persist long enough to enjoy any benefits of a sacrifice. If we focus on larger sub-personal units, we will always have continuum many overlapping person segments at a given time, and it becomes indeterminate which one is engaged in comparing sacrifices and benefits. These problems disappear if the relevant unit is the maximal set of continuous person segments. This makes sense, given that many values, beliefs, desires, and intentions that are distinctive of persons pertain to our lives as a whole rather than segments of it. Brink follows Korsgaard's claim (to be discussed in 3.2.1) that working towards these ends requires close cooperation amongst our person stages. They are thereby unified in a sense that makes it legitimate to focus on the person as a whole.

An important question is how an agent can be 'all there' if some of his parts lie in the future. In response, Brink draws an analogy to nations.²⁰⁷ They are represented by and act through some of their parts, namely representatives. We do not therefore conclude that nations cannot act or possess interests. Similarly, the present self can act as a representative of the temporally dispersed entity, the person. If so, the fact that the temporally extended person is not 'all there' is not a reason to deny that its interests determine what agents have reason to do.

²⁰⁶Brink 1997a, pp. 110-6; Brink 2017, sect. 3.

²⁰⁷Brink 1997a, pp. 115-6.

Another worry stems from fission. It suggests that identity is not necessary for securing what matters. Identity with the beneficiary thus does not seem necessary for justifying a benefactor's sacrifice. Why, then, should concern have an agent bias, as *prudence* requires? Brink suggests two replies.²⁰⁸

First, psychological continuity can extend a person's interests. Even if my fission products are distinct from me, psychological continuity extends *my* interests such that their benefits compensate *me* for a sacrifice *I* make now. This move modifies the nature and range of my interests, but would keep *prudence* intact.

Second, we could introduce *quasi-prudence* which requires benefactor and beneficiary to be psychologically continuous rather than identical. Personal identity and compensation are sufficient for justifying *prudence*. But *quasi-prudence* is sensible even in the absence of identity, provided there is psychological continuity.

Like Whiting, Brink thinks that because intimates share and mutually shape experiences, beliefs, desires and actions, they are justified in regarding their individual interests as extended by the interests of the other.²⁰⁹ Fission shows that psychological continuity is what extends a present self's interests forward in time. And since psychological continuity can obtain interpersonally, interests can also be extended across persons. Fission is one instance of such extension.²¹⁰ But there are more mundane

²⁰⁸Brink 2017, sect. 4. See also Brink 1997a, pp. 125-8.

²⁰⁹Brink 1997b, p. 126; *ibid.*, p. 141.

²¹⁰If fission preserves what matters, it is normally understood to be as good as ordinary survival. If the fissioner's interests extend to fission products, one counterintuitive consequence is that fission can be *better* than ordinary survival: since there are two fission products, we stand a chance to receive compensation *twice*. In principle, Brink could argue that in such cases, the net value or utility per compensation is

ones, too. Intimate friends share their mental lives to a significant degree; their individual experiences, beliefs, desires, ideals and actions depend upon those of the other. In contrast to fissioner and fission products, the shared mental lives of friends are shaped reciprocally rather than one-directionally. In this sense, Brink thinks that the closest analogue to fission is the parent-child relation: the child owes its existence and physical and psychological nature to the parents,²¹¹ just like the fission products owe their characteristics to the fissioner.

Regardless of symmetry and direction, psychological continuity and connectedness extend an agent's interest, and thereby give even rational egoists a "derivative but non-instrumental reason to be concerned about others."²¹² What Brink means is that an agent's consideration of the good of other, psychologically continuous beings is not just a useful means, but valuable for its own sake in virtue of being constitutive of the agent's own good.²¹³ The only reason to care more for oneself than for others is that psychological continuity and connectedness are stronger in the intrapersonal case.

The resulting picture is complex. Brink endorses Sidgwick's idea that the separateness of persons is indispensable for making sense of prudence; personal identity matters in the sense that the separateness of persons is the backdrop against which special concern for one's own future in gen-

lowered by the fact that someone else—the other fission product—receives it, too. If so, fission need not be better than ordinary survival. But the prospects of this reply seem limited. As argued earlier, *some* forms of compensation depend on uniqueness (2.5.3). But it seems doubtful that compensation always does.

²¹¹Brink 1997a, pp. 126-7.

²¹²*Ibid.*, p. 127.

²¹³Brink 1997b, p. 147.

eral, and *prudence* in particular, make sense.²¹⁴ This being said, personal identity between benefactor and beneficiary is not necessary for making a sacrifice reasonable. Psychological continuity either extends a person's interests beyond the separate unit she is, or grounds a *quasi-prudence* according to which continuity rather than identity matters.

Whiting's discussion suggested even more. It is not just that identity is not necessary for special concern. Brink agrees with Whiting that Sidgwick's claim should be reversed: special concern is part of what constitutes psychological continuity and thereby the separateness of persons. The latter involves acting from a distinctive perspective on the world. Apart from displaying concern for one's own future and trusting one's own previous deliberations even if imperfect and inferior to others' deliberations, this involves having personal desires and intentions.²¹⁵ Moreover, individuals stand in social relationships that help to define their lives and generate reasons to care about others. In this picture, an agent's reasons to care specially about her own future does not *derive* from the separateness of persons. Instead, the agent's bias is part of what *constitutes* her separateness.

Besides reversing the relation between separateness and prudential concern, Brink takes Whiting to show that psychological continuity obtains not only within the life of one and the same person, but also across persons:²¹⁶

²¹⁴Brink 2017, sect. 5.

²¹⁵Brink 1997a, pp. 103-5.

²¹⁶If I understand Whiting correctly, Brink's reading here is not precise. For example, he writes: "Just as special concern is partly constitutive of interpersonal psychological continuity, Whiting argues, special concern for ones future self is partly constitutive of intrapersonal psychological continuity" (2017, sect. 6). In contrast, Whiting does not talk about interpersonal psychological continuity. In her picture,

“[i]nterpersonal psychological continuity is quite common and found in many form of interpersonal association between spouses and domestic partners, family members, friends, colleagues, neighbors, etc. In these cases, the mental states and intentional actions of each person depend on those of her associates in various ways.”²¹⁷

He describes some dimensions of interpersonal psychological continuity. It includes affective components which make us fond of our friends and future selves. It also includes behavioral dispositions to help with their projects, and to invest efforts into their success and well-being. Finally, it involves trusting their advice, help, criticism and support.

In both the intra- and the interpersonal cases, psychological continuity is “anchored”²¹⁸ in a thinner relation that does not require, and is prior to, special concern. In the intrapersonal case, it is something like Whiting’s notion of *thin* psychological continuity. In the interpersonal case, individuals stand in “associative relations”²¹⁹ through past interactions and shared histories. These weaker relations get reinforced through caring and planning for one’s personal desires and potential friends respectively, and thereby give rise to personal identity and friendship. In this way, personal identity depends on special concern for one’s future selves, not the other way round.

What kind of concern is it that partly constitutes personal identity? Special concern requires that one cares differently for oneself than for mere strangers. Brink has shown that *prudence*, i.e. an agent bias in combination with temporal neutrality, is one rationally admissible way

special concern for friends is constitutive of *friendship*, not of psychological continuity. She seems agnostic on whether there is genuine psychological continuity across persons; cf. Whiting 1986, pp. 557-8.

²¹⁷Brink 2017, sect. 6.

²¹⁸*Ibid.*, sect. 6.

²¹⁹Brink 1997b, p. 143.

to do this. But special concern need not take a *prudential* form. Brink seems to agree with Whiting that it is up to us how we weigh personal interests and projects with those of others.²²⁰ The same goes for time bias. Temporal neutrality could be defended as a normative claim, but it is not a constraint on practical reason that arises from concern's constitutive role for personal identity.

One remaining question is how special concern can both require personal identity, more precisely the separateness of persons, *and* be constitutive of personal identity.²²¹ This view unites insights from Sidgwick and Whiting. But it seems threatened by circularity, given that Brink also endorses a *no priority thesis*: neither explanatory direction is more fundamental than the other. He admits that this symmetric dependence looks circular, but proposes to distinguish two different ways in which personal identity and special concern depend on each other: special concern is *metaphysically* prior to personal identity because it contributes to the obtainment of personal identity.²²² But personal identity is *normatively* prior to special concern because special concern can be justified only if there is personal identity. If persons are separate, then there can be intrapersonal compensation that is automatic, which justifies caring about these separate units in a special way.

²²⁰Like Whiting, Brink wonders whether persons could be altruistic enough to renounce special concern for their own future and have self-concern “be regulated by her ultimate aim in which her own good is just one among many, with no privileged position” (2017, sect. 7).

²²¹Further below, I argue that Brink tacitly shifts from concern constituting the separateness of persons to concern constituting personal identity (3.1.3).

²²²Brink (2017, sect. 8) is neutral on the exact constitutive role of special concern. If special concern is distinguished by affective components, then it might be neither necessary for nor exhaustive of the psychological continuity relation that constitutes personal identity. If special also includes dispositions to trust, cooperate, plan, and invest, then it is arguably necessary for personal identity.

3.1.3 Reflections: The Constitutive Role Of Attitudes

Each of these positions offers an important insight about how personal identity relates to future-directed concern. Whiting argues that special concern for our friends is justified. Thus, special concern for future selves can in principle be justified as well, even if the numerical distinctness objection against the complex view goes through. Brink explains how we can make sense of the finding that special concern presupposes the separateness of persons, while personal identity itself depends on our special concern for our future selves. Special concern depends *normatively* on the separateness of persons; only if persons are separate is special concern sensible. Personal identity depends *metaphysically* on special concern; the latter partly constitutes personal identity.

It is easy to overstate Whiting's analogy. Earlier, I mentioned Parfit's worries about assimilating concern for our own futures to concern for others (2.5.2). He thinks that the analogy

“still leaves something out. Perhaps some people care about themselves only in the kind of way in which they care about certain other people. But, when we think about our futures, most of us have another attitude, which might be called [...] instinctive concern. Such concern has [...] a ‘special and urgent quality’. This concern may not be stronger than our concern for certain other people, such as our own children. But it seems distinctive.”²²³

I argued that Parfit and others have not made a convincing case that we have concern with this distinctive quality only for ourselves. But even if one were convinced that concern for ourselves feels fundamentally different from concern for others, Whiting's view is not in tension with this.

²²³Parfit 2007, pp. 21-22.

She is not making claims about what concern for our future selves *is like*. Her analogy pertains to the *justification* of the respective concerns, not their *quality*. In fact, Whiting's picture offers resources to make sense of a special and urgent quality of concern for our own futures: prior to coming to care about our future selves, we are already *thinly* psychologically continuous with them. The fact that there is this antecedent connection to our future selves that gets reinforced by coming to care about them could be a first step towards explaining how once concern for them has developed, there is an urgent immediacy to its quality that is absent when caring for others.²²⁴ From here, one could argue further that such antecedent connection is lacking for others, and thereby explain how concern for oneself is special. But on the other hand, one could point out that prior to becoming friends, there is some interrelatedness between us and others that is the counterpart to *thin* psychological continuity. It could be a general concern for others which we have regardless of whether they are our friends, and which then gets reinforced and supplemented through acquaintance and friendship. Or it could be the initial shared experiences and interests that precede the development of friendship. Either way, we would have resources for expressing that concern for our own futures is not fundamentally special and/or different in kind from concern for others. The fact that Whiting's analogy can be developed further in these directions shows that she is not committing

²²⁴Recall Velleman's claim that the immediacy of my concern for my own future requires that I regard the future self unselfconsciously as me—something which I supposedly cannot do for my fission offshoots. With Whiting, we could offer an alternative, less demanding explanation: it is not identity that must be presupposed; instead, being *thinly* psychologically continuous is what accounts for and enables the concern's immediacy.

herself to the particular picture that Parfit deems shorthanded.

For the following reason, Whiting's analogy might not be a decisive reason to reject the extreme claim. While it is difficult to be entirely without friends, we do renounce friendships and thereby, according to Whiting, our second-order concerns for the first-order desires of the former friend. Similarly, while it is difficult, if not incompatible with being a person, to renounce the totality of our personal desires and maintain exclusively impersonal desires, we do sometimes renounce *some* of our personal desires. This being said, Whiting further maintained that it is *rationally admissible* to be without personal desires just as it is rationally admissible to be without friends. Here, one might be worried that Whiting begs the question against Butler and other proponents of the extreme claim. It was precisely their worry that concern for our futures becomes unmotivated. For them, the insight that there is no rational requirement to have friends will turn the analogy on its head—surely we do have a rational requirement or at least a strong motivation for future-directed concern. Thus, friends and future selves are *disanalogous* after all.

In response, Whiting could insist that personal desires are not rationally required, but accept that personal desires are indispensable in a broader sense, e.g., essential to our flourishing. Or she could admit that the analogy between friends and future selves can go both ways. Maybe concern for friends and future selves *is* rationally required. Further arguments would then be needed from Butler and others to explain why there is this rational requirement. Until then, the analogy, if received charitably, suggest that denying this rational requirement would not be absurd and discontinuous with other practices that involve con-

3.1. THE IMPORTANCE OF THE MORE PARTICULAR FACTS 97

cern. Thus, Whiting is indeed in disagreement with Butler and others who think that concern is rationally required. But she is not just begging the question. She is offering an argument from analogy against the requirement. And if the simple view struggles just as much with giving reasons for concern for our futures (2.4.1), Butler and others actually take an interest in accepting the argument. As a result, we are *prima facie* entitled to retain psychological continuity theory even if it assimilates the justification of concern for our own futures to concern for our friends.

Brink and Whiting agree that a momentary person stage cannot be considered a full-fledged agent. Amongst others, the possession, implementation, and satisfaction of intentions and desires requires a temporal extension which person stages lack. The locus of agency is the person whom the person stage represents at a time. Brink sees person stages as representatives of a whole person just like some individuals are representatives of states. Whiting and Brink compare the way in which we care about our future selves to the way we care about our friends. Once friendship is established, I maintain a second-order desire for the first-order desires of my friends. Their good is a constitutive of my own good. Similarly, the good of my present self makes essential reference to the good of my future selves.

They further share the idea that particular attitudes I take towards myself generate further, more general reasons for future-directed concern. In Whiting's case, it is the possession of primitive concerns in the form of personal desires. Brink takes on much of this picture and argues that personal desires give rise to an agent bias that is part of what constitutes

the separateness of persons, a necessary condition for compensation being *automatic*. Neither proposal faces a deep circularity threat. It is true that part of the justification of special concern for our own futures mentions the fact that we are somehow caring for us in the first place. But this seems unproblematic. First, initial concern or practical necessities get us going in caring about ourselves in a more general way. Once this has happened, *new* reasons for concern arise, for example through the fact that the general concern contributes to the individual's separateness from others. Initial concern and the resulting general concern are not responsive to the same kinds of reasons.

Does either the initial or the resulting general concern presuppose personal identity? On why we should care specially for ourselves, Brink says: because if your future selves receive benefits, your compensation is *automatic*. On reflection, this reasoning raises questions. He does not say much about what it means for compensation to be automatic, and even less about why this is important. On the face of it, one would expect that what matters for the rational egoist is that she gets compensated, whether automatically or not. I suppose that automatic compensation is *maximally reliable*: the agent can be *sure* that *she* will be compensated if benefits are accrued intrapersonally, i.e. the future self receiving benefits is psychologically continuous and connected with the agent. If the benefit goes to someone else, it is not clear whether the agent feels compensated. However, note that by Brink's own lights, benefits to someone else can compensate me. One of his key contributions is that psychological continuity and connectedness can obtain across distinct persons. It is not clear why in such a case, compensation should be less automatic.

3.1. THE IMPORTANCE OF THE MORE PARTICULAR FACTS 99

Granted, the agent might feel compensated only if the beneficiary is sufficiently connected and continuous with her, such that the agent's interests are extended the beneficiary. This seems to be a given for the agent's future selves (although see 7.1). But the agent cannot be sure that she is continuous and connected in this way if the recipient is *someone else*. If so, this suggests two things. First, automaticity of compensation matters only derivatively, in virtue of being sufficient for a future benefit being *reliably* compensating in the eyes of the present agent. Secondly, automaticity was key in justifying the agent bias. But now it seems that the justifiedness of the agent bias is relative to the agent's epistemic standing: it depends on whether or not the agent has good reason to think that the future benefit compensates her. Automaticity gives one such reason. However, were she absolutely sure that a future benefit to someone else compensates *her*, she would have no reason for an agent bias in this instance. Thus, whether an agent bias is justified is contingent upon the agent's evidence and interrelatedness with others. The more reason she has to think that benefits to others compensate her, the less an agent bias seems motivated.

All this need not be a problem for Brink's project, which is about rational *admissibility* in the first place: the egoist can justify caring for others. And she can justify having an agent bias but no time bias. It is not an objection to Brink that we *need not* be rational egoists. Moreover, if I understand correctly, there is a sense in which interpersonal compensation actually reduces to intrapersonal compensation. The reason why benefits to others can compensate us is that our interests are extended to them. Their good is part of our own good. In this sense,

we are still receiving the benefit intrapersonally, but via the other as a vicarious, interpersonal beneficiary. Somewhat in contrast with Brink's own rhetoric, this would mean that compensation is necessarily intrapersonal, and talk about the significance of intrapersonal compensation's automaticity misleading; intrapersonal compensation is the *only* kind of compensation there is.

This leads to the question in what sense persons are separate units, and thereby potential recipients of intrapersonal compensation. Referring to Sidgwick, Brink says compensation requires the distinction between persons to be real and fundamental. He calls this realism about personal identity. This sounds as if metaphysics is crucial for future-directed concern after all. The latter cannot be reasonable if metaphysics fails to establish the separateness of persons. Against this, and in the spirit of (2.4.2), it seems to me that the needed notion of separateness is metaphysically much more austere than Brink and Sidgwick make it sound. Even in paradigmatically anti-realist views, separateness could be preserved. For example, even if persons are just fictional constructs out of sequences of experiences, we can maintain a sense in which it is one and the same constructed self that takes an interest in receiving benefits for its earlier sacrifice. In terms of Velleman's position (2.5.2), this requires that there is a future self whom the agent can now regard as *her*. But this requirement seems compatible with a wide range of metaphysics, including anti-realist ones. I am not suggesting that the agent can make arbitrary choices about which future self she considers *her*, only that the requirement is silent on *what it takes* to be *her*. We can thus conclude that the reasonableness of future-directed concern is

3.1. THE IMPORTANCE OF THE MORE PARTICULAR FACTS¹⁰¹

constrained by metaphysics in the sense that a separate unit is needed in which sacrifices and benefits are balanced, but that the constraint itself is minimal. This deflates Brink's claim that special concern for our own futures presupposes personal identity.

Still, Whiting and Brink provide detailed resources for strengthening the case against the extreme claim further. Special concern for our own futures can be justified within the complex view. Its justification mentions personal identity and separateness in a sense that can be fully captured by accounts of personal identity in terms of relations of continuity and connectedness. Moreover, what matters obtains within the life of persisting persons, but can also be realized *across* persons. And the relations that ground special concern are not straightforwardly constrained by person boundaries. They even play a part in defining what the person boundaries are.

Finally, one striking and underexplored question is how essential *psychological* continuity is to these positions. Consider an analogy not between friends and future selves, but between biological relatives and future selves. Interestingly, this substitution preserves a range of aspects of Whiting's initial analogy: the relation between others and myself is similar in kind to the relation between my present and future self. The analogy lends itself to the rhetoric of *primitive* concerns for biologically related selves giving rise to *general* concern for their well-being. There is even a remote and admittedly complicated sense in which biological continuity has concern as a component: caring in certain ways is adaptive and brings about facts of biological relatedness—just like caring in certain ways shapes future strands of psychological continuity. Turning

to Brink, surely biological continuity is one candidate relation for underpinning the separateness of persons, and for extending an individual's interest forward in time. On the face of it, biological continuity can play a similar role as psychological continuity does for Whiting and Brink.

On the other hand, the analogy breaks down at certain points. Biological continuity might sometimes provide *prima facie* reasons for concern, interpersonal compensation, etc. But the reasonableness of these attitudes is easily defeasible.²²⁵ I am neither particularly concerned nor compensated by benefits to a close relative whom I have never met, or whom I have met but whose character and values deviate radically from mine. And it is conceivable to be perfectly justified in caring and feeling compensated despite biological discontinuity. Fancy cases include mentality transfers between biologically discontinuous bodies (4.2), more mundane examples are, e.g., step-parents and -child relations. These reservations about the connection between biological continuity and what matters suggest that the prospects for biological analogues to Whiting's and Brink's positions are dim.

3.2 Self-Constitution Accounts

According to the foregoing positions, psychological continuity and connectedness through intentions, desires, and other attitudes ground the reasonableness of concern. However, one might worry that these relations still fall short of capturing what is important and valuable about our future experiences. The following positions suggest that we need

²²⁵See also D. Shoemaker 2007, pp. 321-2.

to look beyond the numerical identity of persons over time, and instead focus on other senses of identity.

3.2.1 Korsgaard On The Unity Of Agency

Whiting suggested that in order to see why the extreme claim is mistaken, we need to realize that there is more to psychological continuity than just memory. Korsgaard takes precisely this thought as a starting point. Via a different route, she arrives at conclusions that seem congenial to Whiting's picture.

Parfit's account of personal identity focuses on the experiences and memories of person stages. He argues that they fail to constitute a deep further fact. In response, Korsgaard stresses that persons are agents. Agents are not just passive bearers of experiences. They are able to reason and deliberate about how to act, and can act in accordance with their deliberations. Parfit's account focuses only on experiences, but does not mention agency. His account is thus incomplete.

Korsgaard claims that even if the simple view is false, I have *practical* reasons for regarding myself as the very same rational agent over time.²²⁶ She motivates this diachronic case by means of considerations about the synchronic case. At a given time, I see myself as one particular person not because of metaphysical claims I accept, or a unity of which I am conscious. Instead, two kinds of activities unify me at a time. First, I eliminate conflicts among my various motives. Some of my motives are incompatible with each other, and through reasoning I determine which motives I adopt and implement. Secondly, I do not merely act according

²²⁶Korsgaard 1989, p. 109.

to my strongest desires. Instead, I deliberate about conflicts among my desires. I choose on which ones I act based on reasons. Such choice is expressive of my will.

“When you deliberate, it is as if there were something over and above all your desires, something that is you, and that chooses which one to act on.”²²⁷

Being “a law to yourself”²²⁸ in this sense does not presuppose the truth of the simple view, i.e. the existence of a separate or irreducible self. The unity of agents arises from the practical necessity of making deliberative choices.

Against the practical conclusions which Parfit draws from the complex view, Korsgaard notes that the role of metaphysics in this deliberation is not settled. Metaphysical findings *can* have impacts on my choices. For example, it can be significant to learn that a future self will be me. But whether and which impact metaphysics has on my choices is not determined prior to my own, personal deliberation.

“It is practical reason that requires me to construct an identity for myself; whether metaphysics is to guide me in this is or not is an open question.”²²⁹

These are considerations about the synchronic unity of agents. Korsgaard transfers these insights to the diachronic case. In one sense, her view is consistent with Parfit’s: personal identity is not a deep further fact, but rather like the identity of a nation. But Korsgaard clarifies that a state provides the better analogy than a nation. The connection between present and future selves is not as loose as the one between people

²²⁷Korsgaard 1989, p. 111.

²²⁸*Ibid.*

²²⁹*Ibid.*, p. 112.

who happen to live in the same territory and share some history. A state is the result of the citizens' decision to constitute a single unified agent. Persons over time are similarly unified. They do not merely consist in person stages that happen to precede and follow each other. Their decision to unify by extending their interests to each other is a practical necessity; there is an imperative to unify—just like for a Kantian, there is an imperative to form a state.

This practical necessity to unify arises from the fact that agents make choices whose fulfillment takes up time, e.g, choices about careers, health, friendships, family, life plans. Even a trivial choice involves a future-directed attitude. This is why for Korsgaard, the question whether and why a person stage should be concerned about a future self is misleading. The question suggests that the stage already has independent reasons for action, and needs to be convinced why it should take future stages into account. But more appropriately, the stage's reasons derive from identification with future stages in the first place.

“When the person is viewed as an agent, no clear content can be given to the idea of a merely present self.”²³⁰

It is misleading to demand reasons why one person stage should be concerned about another. Person stages at different times are not disconnected, but closely interrelated through intentions, desires, and other states and attitudes with essentially diachronic orientations. Like Whiting and Brink, Korsgaard thinks that person stages in isolation cannot be full-fledged agents in their own right. They do not have independent interests that could be weighed against interests of other person stages.

²³⁰*Ibid.*, p. 113.

The extreme claim is thus fundamentally misguided.²³¹

Moreover, while she shares Parfit's conviction that interpersonal and intrapersonal concern are not fundamentally different, her reasoning is distinctive: our being unified into agents is what generates reasons that transcend the boundaries of persons. The reasons that arise from unification into agents are called *personal* reasons.²³² And Korsgaard is convinced that each of us is unified not only into the agent who occupies a particular body. Instead, we are also part of and intertwined with other agencies of fellow persons, families, organisations, nations, etc.:

“the personal concern which begins with one's life in a particular body finds its place in ever-widening spheres of agency and enterprise, developing finally into a personal concern for the impersonal—a concern, that is to say, for the fate of one's fellow creatures, considered merely as such.”²³³

If I understand correctly, not only does the synchronic case generalize to the diachronic case; the diachronic case generalizes to the interpersonal case. Person stages at a time are so intertwined with person stages at other times that their interests cannot be understood in isolation. Agents “come in different sizes”²³⁴, and persons are only the smallest units. The agents occupying our bodies are not only unified; they further unify into interpersonal agents. Unification into a group agent gives individual agents *personal* concern for others. Personal reasons have force not just

²³¹Korsgaard 1989, pp. 126-7.

²³²Note that this notion of a *personal* reason or concern is not exactly the same as what Whiting calls a *personal* desire. For example, Korsgaard's personal reasons are the result of a practical necessity to unify, whereas Whiting makes no such claims, and seems to allow that person desires can be renounced entirely. And unlike Whiting's notion, Korsgaard's personal reasons can be essentially tied to *others*, not the agent; see below.

²³³Korsgaard 1989, p. 127.

²³⁴*Ibid.*

for an agent's own life, but also govern her relationships and interactions.

Despite a number of disagreements with Parfit, Korsgaard's view seems to have the potential to invite improvements rather than radical revisions of psychological continuity theory. She suggests that psychological continuity and connectedness should encompass more than passive mental states of memory and experience. Instead, they should be extended to active states and attitudes that are constitutive of agents. Under this description, Korsgaard's and Parfit's views have structural similarities; they just pick out (or emphasize) different psychological relations. Another parallel is that what matters can come apart from identity. For Korsgaard, the unification of agents is what generates practical reasons. Her phrasing of agents occupying particular bodies²³⁵ suggests that it is possible that agents come apart from their bodies. Presumably, agency need not adhere to the logic of identity, e.g., it can branch.

What is distinctive and goes beyond psychological continuity theory is Korsgaard's normative claim that in conceiving of oneself as an agent, one thereby incurs the practical necessities she describes. This motivates practical conclusions that differ significantly from Parfit's. For example, Korsgaard rejects paternalistic interventions and utilitarian distribution patterns on this basis.²³⁶ Although Korsgaard is not explicit about this, I also suppose that since an agent is unified over time, she has no reason to accept a Parfitian discount rate for future-directed concern²³⁷ that reflects the degree of psychological connectedness. However, these disputes do not seem to rest on a fundamental difference in descriptive pictures;

²³⁵For example *ibid.*, p. 126.

²³⁶*Ibid.*, Ch. V.

²³⁷Parfit 1984, pp. 313-4.

Korsgaard’s diachronic relations between person stages seem in principle describable in terms of rich enough variants of psychological continuity and connectedness.

Then again, some of Korsgaard’s claims are ambiguous. When she says that, as quoted, practical reason requires me to construct an identity for myself, it is not clear what kind of identity is being constructed here. It could be that exercising practical reason generates, or is constitutive of, a person’s numerical identity over time. But she could just as well be read as saying that regardless of which facts about numerical identity over time obtain, agency and practical reason provide the necessary and practically relevant unity to justify judgements and attitudes about moral responsibility and future-directed concern. This reading seems appropriate when she says that understanding ourselves as agents follows from seeing ourselves as first causes of our actions, an attitude that is

“forced upon us by the necessity of making choices, regardless of the theoretical or metaphysical facts.”²³⁸

In other words, it is not entirely clear whether her position is intended to capture personal identity over time, or merely the sameness of a practically relevant unit that can come apart from personal identity.²³⁹ I argue below (3.2.3) that if Korsgaard intends to focus on the sameness of a practically relevant unit, the question arises why she takes issue with

²³⁸Korsgaard 1989, p. 120.

²³⁹DeGrazia (2005, p. 79) writes: “I cannot tell whether [Korsgaard] is addressing numerical identity, narrative identity, or some merging of the two”. Mackenzie (2008, pp. 2, 8-11) understands her as discussing the “practical identity” of agents, which must presuppose our numerical identity or unity over time and is constrained by facts about embodiment, but arises from practical rather than metaphysical necessity. Schechtman (2014, pp. 44-9) reads her as speaking primarily to a form of agent identity that is distinct and independent from personal identity and other theoretical or metaphysical notions.

Parfit's reductionism. After all, the latter position is a proposal about the numerical identity of persons over time. By itself, it is neutral on whether persons are the practically relevant unit.

I now turn to Schechtman's position which draws a clear boundary between the numerical identity of persons over time and a more practical sense of identity.

3.2.2 Schechtman On Narrative Identity

Schechtman points out that there are two senses of personal identity that are usually conflated.²⁴⁰ *Reidentification* criteria address

“the question of what makes a person at a time t_2 the same person as a person at the time t_1 ”²⁴¹.

Schechtman emphasizes that there is another important question about persons and their existence over time, the *characterization* question:

“this question asks which actions, experiences, beliefs, values, desires, character traits, and so on (hereafter abbreviated “characteristics”) are to be attributed to a given person. [...] [C]haracterization theorists ask what it means to say that a particular characteristic is of a given person.”²⁴²

In contrast to the reidentification question, the relata mentioned in the characterization question are not two person stages, but one person and one or more characteristic. The attribution of characteristics comes in degrees. Some characteristics are part of us and our lives, but only incidentally so. They are part of our histories, but they are not central

²⁴⁰For endorsements and applications of Schechtman's view, see for example Blustein 1999, and DeGrazia 2005, ch. 3.

²⁴¹Schechtman 1996, pp. 1-2.

²⁴²*Ibid.*, p. 73.

to who we are. This is different for certain other characteristics. They determine our identity in the sense of defining us and making us the individuals we are. Here, characterization is

“not [...] about ‘identity’ understood as ‘the relation which every object bears to itself and nothing else,’ but rather as ‘the set of characteristics that make a person who she is’.”²⁴³

The reidentification and characterization questions are related. In order for an action to be attributable to a person, we need reidentification identity between the addressee of the attribution and the agent who performed the action. Still, Schechtman maintains that the questions are distinct.

Schechtman thinks that in the past, philosophers tried to provide criteria for personal identity that manage to justify what Schechtman calls *four basic features* of our lives: we survive events, we ascribe moral responsibility, we have self-concern, and people who carry burdens have a right to be compensated. Psychological continuity is the most promising relation for making sense of the four features. However, psychological continuity theory has a number of shortcomings. First, direct psychological connections between person stages are sufficient for personal identity. But as Reid’s *brave officer* objection²⁴⁴ shows, they are not necessary. Psychological continuity theorists are forced to concede that indirect connections of psychological continuity can suffice. Schechtman claims that this move deprives the view of its intuitive appeal.²⁴⁵ Direct psychological relations did seem to capture a good sense in which another person

²⁴³Schechtman 1996, pp. 75-6.

²⁴⁴Reid 1785a, p. 276.

²⁴⁵Schechtman 1996, pp. 28-30.

stage's experience are mine. In contrast, indirect psychological connections do not seem as important as we thought identity would be.

Secondly, identity is all-or-nothing, psychological connectedness allows for degrees. Parfit and others thus introduce a threshold of psychological connectedness that is necessary for personal identity. Schechtman thinks that any threshold is inevitably arbitrary.²⁴⁶

Thirdly, psychological continuity theory takes personal identity to consist in *similarity* relations between person stages. But similarity is unsuitable for grounding the four features.²⁴⁷ I am responsible only for what *I* did, not what someone who is very similar but distinct from me did. I am specially concerned about what *I* experience, not what someone who is just like me but distinct from me experiences. Psychological continuity thus leaves open why I should care about my future selves.²⁴⁸ In other words, Schechtman endorses the extreme claim, and her motivations are similar to the ones described earlier (2.1).

The key to address these shortcomings is to realize that reidentification is unsuitable for grounding the *four features*. Instead, characterization questions need to be discussed. Schechtman's proposal for answering the characterization question is called the *narrative self-constitution* view. Persons do not only conceive of themselves as a persisting subject that has past, present and future experiences. They see these experiences as parts of a life story that partly constitutes their self-conception.²⁴⁹

²⁴⁶[Ibid.](#), p. 43.

²⁴⁷[Ibid.](#), p. 52.

²⁴⁸[Ibid.](#)

²⁴⁹According to Mackenzie (2008, p. 9), this is a parallel to Korsgaard's view: both think that self-constitution is what brings about practically relevant forms of identity.

“a person creates his identity by forming an autobiographical narrative—a story of his life. [...] According to the narrative self-constitution view, the difference between persons and other individuals [...] lies in how they organize their experience, and hence their lives. [...] 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 (in the sense at issue in the characterization question) 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.”²⁵⁰

There are constraints on what qualifies as an admissible narrative: it should not contain obvious falsities, and it shall cohere reasonably well with the narrative others tell about oneself. The reason is that personhood is a social concept; one essential component of being a person is playing one’s part in a web of social relations. Amongst others, this requires “a self-concept that is basically in synch with the view of one held by others.”²⁵¹ Moreover, the episodes of the narrative need to fit and connect reasonably well such that they make up an intelligible story.²⁵² Most persons will have their narratives only subconsciously or implicitly, and this is compatible with Schechtman’s position.²⁵³ However, it should be in principle articulable.²⁵⁴

The crucial feature of proper narratives is that they extend a person’s consciousness over time. Consciousness *at a time* is unified. Synchronic experiences are not a mere set or conglomerate of distinct component experiences. Instead, different sensual inputs, such as visual, auditory, tactile and olfactory inputs

²⁵⁰Schechtman 1996, pp. 93-4.

²⁵¹*Ibid.*, p. 95.

²⁵²*Ibid.*, pp. 96-7.

²⁵³*Ibid.*, p. 105.

²⁵⁴*Ibid.*, p. 114.

“mutually interact in a way that produces a unified experience in which each plays a role, and plays the role it does because it serves as an interpretive context for the others, which in turn provide its context.”²⁵⁵

Narratives achieve something similar in the diachronic case. Just like mutual interaction among simultaneous inputs and states results in a unified experience at a time, narrative interconnectedness among a person’s moments of consciousness at different times results in them being unified over time. The individual comes to understand herself as temporally extended rather than momentary.

“The moments of conscious awareness in a single person’s life are not distinct entities that are somehow strung together, but rather a dynamic interactive system that integrates to produce a subjectivity that extends over time. [...] [T]he experiences woven together into a person’s narrative interact and alter one another in such a way that the narrative itself becomes the primary unit. [...] The experience of a person is thus had by an extended narrative subject, and not by a time-slice.”²⁵⁶

In other words, having a narrative structure unifies the moments of experience and thereby produces one single, persisting subject.²⁵⁷

In contrast, children, dementia patients, and others cannot organize their experiences into a narrative self-conception.²⁵⁸ For them, experiences remain without context and do not hang together with other mo-

²⁵⁵*Ibid.*, p. 139.

²⁵⁶*Ibid.*, pp. 143-4.

²⁵⁷Schechtman offers analogies for how her picture of the relation between a person’s life and her experiences contrasts with others. Sameness of substance views (of which the simple view is an instance) see the person’s experiences like beads on a string: distinct, separable experiences belong to one and the same person. The complex view compares it to a river: a continuous flow of distinct, separable experiences is all there is to a person’s life. In contrast, the narrative self-constitution view offers the picture of “a complicated stew or soup” in which the different ingredients interrelate and combine into something new by “seasoning and altering one another” (*ibid.*, p. 144).

²⁵⁸*Ibid.*, pp. 146-7.

ments of consciousness. Instead of being extended in time, they experience only a succession of disconnected episodes. They are unable to project themselves forward or backward in time. Consequently, they lack the phenomenology and capacities of narrating subjects, and cannot satisfy the conditions of personhood that are relevant for *characterization* and underpin the *four features*.²⁵⁹

The narrative self-constitution view promises that

“the kinds of connections between the different temporal portions of a person’s life which provide a basis for the four features are *created* by the having of a self-conception.”²⁶⁰

How does having a narrative self-conception ground the four features? I here describe the example of future-directed concern. For Schechtman, being a narrator has two crucial aspects. First, anticipating a future experience in one’s narrative influences the character of present experiences in a certain way:

“what makes a future one’s own on this view is that its anticipated character can cause pleasure or pain in the present. The present person thus has an obvious and immediate stake in anticipating a pleasant future.”²⁶¹

Second, the narrative extends the narrator’s consciousness forward.

“Thought of one’s future [...] brings the sense that there is more than just what is present at the moment, and this makes the present a wholly different kind of thing.”²⁶²

By narrating, the individual orients itself towards the future. She considers herself as the subject not only of present moments of anticipation and

²⁵⁹Schechtman 1996, p. 94.

²⁶⁰*Ibid.*, p. 101.

²⁶¹*Ibid.*, pp. 154-5.

²⁶²*Ibid.*, p. 156.

fear, but also of future experiences that are anticipated. Because anticipated experiences shape the quality of present experiences, and because her narrative extends the individual into the future and makes it the case that *she* will be having the anticipated experiences, narrative self-constitution can ground the reasonableness of future-directed concern. The view “capture[s] the intuitive relation between personal identity [in the *characterization* sense; P.H.] and the four features.”²⁶³

Schechtman’s position motivates a suggestion for why it is so hard to decide whether animalism or the psychological continuity theory provides the right criterion of personal identity. She argues that for *reidentification*, animalism is superior, whereas psychological relations are superior for addressing *characterization* and giving reasons for the *four features*.²⁶⁴ The failure to distinguish these question explains why our intuitions conflict. Both criteria are indeed appealing, but each of them with respect to a different question.

Schechtman’s recent work develops several aspects of the narrative self-constitution view further.²⁶⁵ She nowadays thinks that *characterization* is too coarse-grained.²⁶⁶ It does not reflect the fact that some features and actions can be attributed to us qua *moral selves*, whereas others belong to us qua *forensic units*. If I trip and break a lamp, this event is attributable to me qua *forensic unit*: I am a suitable target for the attribution. However, mere attribution does not settle whether I am morally responsible. If I can rightfully be held accountable for the

²⁶³[Ibid.](#), p. 99.

²⁶⁴[Ibid.](#), pp. 68-70.

²⁶⁵It is not always clear whether she intends to *replace* or *extend* her earlier theory; see for example D. Shoemaker [2016](#), p. 325.

²⁶⁶Schechtman [2014](#), pp. 41, 101-2.

action, then it is attributable to me qua *moral self*. This distinction matters because neglecting it has resulted in crude pictures of the relation between personal identity and practical concerns where the former depends straightforwardly on the latter, i.e. x and y are identical iff they are the same moral self.²⁶⁷ Against this, Schechtman proposes that we tie personal identity to forensic unity, i.e. regard persons as *loci* for attributions of responsibility, concern,²⁶⁸ compensation, etc. Her new positive picture of personhood is the *person life view*:

“To be a person is to live a ‘person life’; persons are individuated by individuating person lives; and the duration of a single person is determined by the duration of a single person life.”²⁶⁹

What exactly counts as a person and the protagonist of a person live depends on society’s “person-space”²⁷⁰—the culture, norms, and institutions surrounding person-related practices. I here see the *person life view* as picking up on the idea that the narratives of persons are inherently social concepts: leading a person life requires standing in social relations and interacting with others. Person lives typically involve developmental trajectories, including the possession of different abilities at different life stages, e.g., reflective self-consciousness and self-narration as an adult, but less demanding abilities at early and late stages of life. As such, beings who according to the narrative self-constitution view would not count as narrators can nevertheless live person lives. Schechtman

²⁶⁷Schechtman thinks Locke has mistakenly been read this way.

²⁶⁸We might worry that this distinction works for some practical concerns better than others. While plausible for moral and forensic matters, the proposal to regard persons as *loci* of attribution threatens to overgenerate for concern: it cannot be that I am identical to all those future selves who are plausible addressees for my special future-directed concern.

²⁶⁹Schechtman 2014, p. 110.

²⁷⁰*Ibid.*, p. 114.

argues that as a consequence, personal identity as understood by the person life view is a cluster concept.²⁷¹ Its obtainment is typically a matter of biological, psychological and social relations obtaining and being integrated. But none of these relations is individually necessary and/or sufficient. Any thought experiment in which these relations come apart “is in fact a degenerate case of the more basic relation that contains all three.”²⁷²

3.2.3 Reflections: Self-Constitution And Psychological Continuity

Korsgaard and Schechtman differ in their exact responses to the extreme claim. For Korsgaard, the question why my present self should have concern for my future self is misguided, given the practical necessity to unify. In contrast, Schechtman *accepts* the extreme claim against the complex view as a *reidentification* criterion. I think both can agree that the complex view, by mentioning only facts about distinct person stages, isolated experiences, and continuity relations, cannot provide a rational grounding for concern. Instead of focusing on descriptive pictures, we need to consider *activities* we are engaged in. The construction of practical identities, and/or the narration of our life stories make concern for ourselves and others reasonable, and underpin further practices like compensation and moral responsibility. This seems to fit the strategy described earlier that the extreme claim can be resisted because concern can be justified without appealing to metaphysics (2.4.2).

²⁷¹*Ibid.*, ch. 6.

²⁷²*Ibid.*, p. 150.

Moreover, both Korsgaard and Schechtman would deny Whiting's claim that our concerns and other practical activities are constitutive of personal identity in the *reidentification* sense. What *is* being constituted is the unity of an agent or a life story. These need not coincide with the *reidentification* of a person. Earlier, I suggested that Brink can be read in a similar way, as arguing that the separateness of a person is constituted partly by the individual's possession of an agent bias (rather than the other way round), while the relevant kind of separateness is metaphysically neutral or austere, and merely denotes the existence of separate practical units in which sacrifices and benefits should be balanced (3.1.3).

Korsgaard claims that it is a practical necessity to have diachronic attitudes that unify our selves at different times. Just like in Whiting's picture, there is a parallel between the relations among person stages and the relations among persons. For Korsgaard, in both cases there is a practical imperative to unify and coordinate one's ends. For Schechtman, narratives are what unifies a consciousness over time. The fact that what counts as a proper narrative partly depends on how others see the narrator gives a sense in which intrapersonal unity depends on relatedness with others. But unlike Whiting and Korsgaard, the claim is that intrapersonal and social relations are interdependent, not necessarily analogous.

Like Whiting and Brink, Korsgaard thinks that the fact that I am taking a certain attitude towards my future selves is what generates more general reasons for future-directed concern. In Korsgaard's case, it is the possession of interests and desires whose fulfillment takes up time that

require cooperation with the person stages who occupy my body in the future. Since it is a practical necessity for person stages to unify, there can be one and the same agent at different times even if the simple view is false. Nevertheless, one might be concerned exactly how far this reasoning can reach. While it might be a practical necessity for any person stage to identify with certain other person stages, it is not obvious how a person stage is necessitated to unify with *all other* person stages who succeed her in occupying a certain body. Korsgaard maintains that in isolation, we cannot make sense of the reasons and interests of a momentary person stage. Suppose, however, that a person stage unifies with certain other person stages into a person segment with thoroughly hedonistic preferences. This segment has an incentive to free-ride upon the well-being of the person stages that will come to occupy the body she occupies, and to engage in highly enjoyable yet destructive activities. For her, it is an imperative to be oriented towards the near future, but not further. Thus, the scope of the practical necessity to unify seems relative to the substantive reasons and interests of the constituted person segment. It is a leap to conclude from the diachronic nature of motives, intentions, and desires that persons, qua being agents, face a practical necessity to care for their lives as a whole.

Moreover, note that Korsgaard too can allow for a sense in which special concern for one's own future is not rationally required. Given the desires and interests I have, the unification between my person stages might be relatively loose. In particular, their degree of unification might pale relative to how tightly I unify in agencies with people who are close to me. I do not claim that Korsgaard is committed to these results, just

that they seem compatible with what she says. This deflates the sense in which it is a *necessity* to unify with future person stages that occupy my body.

One potential benefit of self-constitution views is that they can advance debates between competing criteria of personal identity. Our intuitions notoriously point in different directions, sometimes favouring a psychological criterion, while other times suggesting a bodily account (4.2). Some authors thus conclude that there is no fact of the matter which criterion is right, or that our concept of personal identity is incoherent (4.6). By distinguishing personal identity from the identity of agents and/or narrative self-constitution, these views can offer an explanation of what goes wrong: our seemingly divergent intuitions speak to different questions. When theorizing and considering puzzle cases about personal identity, we conflate persons with agents, and *reidentification* with *characterization*. No wonder our intuitions are inconstant.

This suggestion is not only diagnostic, but also therapeutic: the hope is that if we discuss these questions individually, clearer pictures of which criterion addresses which question best will emerge. Here, Korsgaard distinguishes further between what agents are *contingently* as opposed to *necessarily*. She prefers

“a physical criterion of identity, but only a conditional one. *Given the technology we have now*, the unit of action is a human body.”²⁷³

But if one day, agents could carry out unified plans and projects *across* human bodies, the latter would cease to be the basic unit of action.

Schechtman agrees that bodily criteria are appealing, but only for *rei-*

²⁷³Korsgaard 1989, p. 115.

identifying persons.²⁷⁴ Psychological relations are clearly more relevant to the *four features*²⁷⁵— although they remain insufficient when not used to address *characterization* questions. DeGrazia is inspired by Schechtman and her distinction between personal and narrative identity. He argues that animalism gives the right account of personal identity, whereas accounts of narrative identity capture how we value psychological continuity.²⁷⁶ These suggestions sound conciliatory: we can keep both bodily and psychological criteria, but each for a slightly different topic.²⁷⁷

At the same time, the narrative view offers something to proponents of the extreme claim who remain convinced that continuity relations are not enough to ground concern.²⁷⁸ It is narrative interconnectedness that needs to relate moments of consciousness at different times in order for them to unify such that actions and features can be attributed to a subject who comes to care for its life as a whole.

At times, especially Schechtman seems more dismissive of psychological continuity theory than necessary. She thinks that a Whiting-style picture belittles the nature of concern for our own futures.

“[T]hese views do not deny that the present person-stage, who is worrying about the future of the person, is a distinct experiencing subject from anybody existing in the future. What this means, however, is that the concern of the present subject for her future self can never be a concern based on anticipation.”²⁷⁹

Even complex positions that take lengths to *respond* to the extreme claim

²⁷⁴Schechtman 1996, p. 67.

²⁷⁵*Ibid.*, pp. 15-6.

²⁷⁶DeGrazia 2005, p. 80.

²⁷⁷These pictures resemble Olson’s distinction between personal identity and *being the same person*, to be discussed in (4.5).

²⁷⁸For an argument that narratives are not sufficient either, see Crone 2017.

²⁷⁹Schechtman 1996, p. 65.

cannot escape the conclusion that anticipatory concern is misguided. Against this, I argued earlier that special concern can still be taken to arise from *quasi*-anticipation (2.5.2). In treating concern for friends and future selves on par, Whiting's picture provides a particularly promising context for quasi-anticipation. This suggests that refined versions of the complex view of *reidentifying* persons might deal with concern better than Schechtman expects.

I mentioned how Korsgaard can be seen as *improving* psychological continuity theory rather than rejecting it. A similar suggestion can be made with regards to Schechtman. One move in the literature is to make a case that even as a *reidentification* theory, the complex view can meet the challenges of Schechtman and the extreme claim.²⁸⁰ But it has not been remarked that narrative interconnectedness between person stages is plausibly seen as an *instance* of psychological continuity and connectedness. If the latter can comprise memories, experiences, beliefs, anticipations, plans, etc., why can we not add narrative attitudes to the mix? They seem more complicated and nuanced than the foregoing attitudes, but not therefore less psychological in nature. For example, it is plausible to see narrative structures as structures *among* psychological states.²⁸¹ The suggestion to enrich the notion of psychological continuity

²⁸⁰See my earlier discussion of how continuity relations justify concern (2.5). Explicitly with regards to Schechtman, Beck (2015, pp. 314-6) defends psychological continuity theory on the grounds that even if we grant that person stages at different times are necessarily distinct, stages of one and the same person are connected by *causal dependence*. Schechtman overlooks that this might be quite relevant towards the *four features*; for example, my future selves deserve compensation for my present efforts partly because they feel aches *as a result* of my present efforts.

²⁸¹Understood this way, narratives reorganize and relate memories, present beliefs and aspirations, and anticipated future experiences. One worry is that this construction threatens to operate on a set of interrelated states that need to be somewhat unified and integrated *before* they can be subject to narrative reorganization. This is

in this way leaves intact Schechtman's claim that *characterization* rather than *reidentification* questions need to be discussed when justifying the *four features*. But once care is taken to focus on the right question, it seems we become less pressed to reject psychological continuity theory. The central lesson from Schechtman would then be twofold: that *characterization* differs from *reidentification*, and that—as pointed out by Korsgaard as well—psychological continuity and connectedness should be understood in sufficiently rich terms, encompassing active, agency- and narration-related states.

3.3 Conclusion

In this chapter, I described two approaches towards the complex view in view of the objections that motivate the extreme claim. The first is to defend and improve the view (3.1). The complex view can understand personal identity as being constituted by relations that matter. Whiting argues that relations quite similar to those which justify concern for our friends obtain between us and our future selves. Brink takes relations which ensure that we are compensated by benefits to a future self as partly constitutive of the separateness of persons. These positions operate with the resources of psychological continuity theory. It is not obvious how similar moves could be available to proponents of bodily variants of the complex view.

According to the second approach, we need to look beyond personal identity and what it consists in (3.2). Korsgaard argues that the practical

a parallel to Zahavi's criticism (2007) that, amongst others, we need to presuppose a sense of self that precedes narrative constitution.

necessity of person stages to constitute a unified agent makes the request of the extreme claim to provide reasons for concern misguided. If seen as agents, persons are always temporally extended, and the idea of a momentary person stage with independent interests and reasons does not make sense.

Schechtman's narrative self-constitution view contributes a distinctive innovation: it captures a person's *identity* in the *characterization* sense. The *reidentification* sense of personal identity is practically unimportant, at least in terms of giving reasons for the *four features*: caring about survival, ascribing moral responsibility, carrying present burdens for the sake of future compensation, and caring about oneself. In contrast, the narratives relevant to the *characterization* questions unify person stages such that the *four features* become reasonable.

These positions point at the need to operate with a sufficiently rich notion of psychological continuity and connectedness that is not limited to passive states like memories and experiences. In order to be relevant to what matters, the *four features*, etc., agency-related continuities and activities, narrative structures amongst psychological states, and a range of concerns and biases need to be considered. The plausibility and scope of psychological continuity theory hinges partly on the extent to which it can subsume and reflect these different aspects of our mental lives.

As a result, we gain further resources to respond to the numerical distinctness charge against the complex view: first, it is far from obvious that numerical identity, e.g., in the *reidentification* sense, is what matters and should be expected to give reasons for concern. Second, not all distinct things are equally unrelated and unintegrated with each

other. Numerical distinctness amongst person stages is unproblematic if the stages are related in the right way. A rich notion of psychological continuity and connectedness can bind distinct person stages together such that practical concerns stay intact.

Along the way, we also encountered challenges. In different ways, the authors claimed that personal identity is constituted by particular activities of ours and/or components of psychological continuity. In Whiting's words, we *make* our future selves. The question arises whether the authors intend—let alone are entitled—to regard the constituted sense of personal identity as the numerical identity of *persons* over time, or as the sameness or identity of a *practically relevant unit* that could in principle come apart from the person that embodies it. Care and clarity are needed to make precise the claim that personal identity is constituted by special concern and other practical activities.

We can agree with Parfit that positions like these are, as he says, *liberating*. There is a wide range of rationally permissible attitudes we can adopt when making our future selves, constructing narratives, and unifying with other person stages that occupy our body. Maybe we can renounce *personal* desires entirely, as Parfit seems to have done when leaving the glass tunnel. But reflection on the different ways in which the characterized positions understand ourselves as defined through practical activity brings out that these different attitudes towards our futures were always available to us. The range of rationally permissible attitudes does not open up because a deep further fact is absent. Endorsements of the complex view are not necessary for an entitlement to shape *personal* projects and desires in the way we envision.

Chapter 4

Transplant Intuitions

Consider the following case by S. Shoemaker:

“suppose that medical science has developed a technique whereby a surgeon can completely remove a person’s brain from his head, examine or operate on it, and then put it back in his skull (re-grafting the nerves, bloodvessels, and so forth) without causing death or permanent injury [...]. One day, to begin our story, a surgeon discovers that an assistant has made a horrible mistake. Two men, a Mr. Brown and a Mr. Robinson, had been operated on for brain tumors, and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown’s brain in Robinson’s head, and Robinson’s brain in Brown’s head. One of these men immediately dies, but the other, the one with Robinson’s body and Brown’s brain, eventually regains consciousness. Let us call the latter ‘Brownson.’”²⁸²

Suppose that by virtue of having Brown’s brain, Brownson has Brown’s psychology, character, personality, etc. Yet, Brownson has Robinson’s body. Is Brownson Robinson, Brown, or neither? Following Olson, I call intuitions on this and similar questions (see 4.1) *transplant intuitions*.

Transplant intuitions play important dialectical roles in debates on personal identity. They are typically understood as supporting psychological continuity theory and opposing animalism.²⁸³ Olson even calls the judgement that one survives as one’s psychological continuer *the* transplant intuition.²⁸⁴ For some, transplant intuitions lead to a paradox:²⁸⁵ these intuitions are strong and imply the falsity of animalism, a view which otherwise seems intriguing and superior to psychological continuity theory. Animalists try to show that their view can resist transplant intuitions.²⁸⁶ Others think transplant intuitions might appear to support

²⁸²S. Shoemaker 1963, p. 23.

²⁸³Parfit 2012, p. 10. Herdova 2016, pp. 315-6.

²⁸⁴Olson 1997, p. 44.

²⁸⁵Garrett 2003, pp. 41-44.

²⁸⁶Olson 1997, 2015a; Snowdon 2014.

a psychological criterion, but in fact support a bodily criterion.²⁸⁷

In this chapter, I argue that there is a deep problem with transplant intuitions that renders them almost useless for choosing between psychological continuity theory and animalism. I claim that all sides in this debate have vastly overestimated the value and importance of transplant intuitions. I accept that transplant intuitions can play a role in getting our thinking about personal identity started. But for the purpose of identifying the criterion for personal identity over time, they have severe limitations.

I begin by describing transplant intuitions (4.1-4.2) before developing the claim that transplant intuitions are bad candidates to track facts of personal identity (4.3). I argue that this motivates a debunking argument against beliefs based on transplant intuitions (4.4). In the final section, I explain why this does not commit me to objectivism (4.5) or skepticism (4.6) about personal identity.

4.1 Locke's *Prince And The Cobbler*

I speak of 'transplant intuitions', but for my purposes, it is not essential that they concern the literal *transplant* of an organ, e.g., the brain. The feature I would like to focus on is the separation of psychological and bodily continuity. In the *Brownson* case, this separation is brought about by the brain transplant: it makes Brownson, who has Robinson's body (minus the brain), psychologically continuous with Brown. But there are other cases which bring about a similar separation of psychological

²⁸⁷Williams 1970, pp. 179-180.

and bodily continuity without involving literal transplants of organs. For example, Locke regards the soul as the locus of a person's consciousness, and imagines:

“[S]hould the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, every one sees he would be the same person with the prince, accountable only for the prince's actions”²⁸⁸.

Locke's quote allows for at least two readings. First, because everyone sees that the prince is now in the cobbler-body, the person in that body would be accountable for the prince's action. Second, because the person in the cobbler-body would now be accountable for the prince's action, the prince is in the cobbler-body.

In order to set the stage for what I think is problematic about intuitions on cases of this kind, let me provide the following preliminary comments on the two readings. If this case is supposed to inform us about the right criterion of personal identity, then the first reading begs the question. Where the prince goes is precisely what is in dispute. While the reading might *illustrate* a certain criterion of personal identity and invite us to consider who will be accountable for what, it goes no way to *answer a question* about personal identity. The reading does not advance our understanding of where the prince goes. In particular, it does not give reasons to think that the prince will inhabit the cobbler-body.

The second reading does speak to the personal identity question. But it is not obvious why the connection between the sentence parts should hold. It presupposes that the norms of accountability track personal

²⁸⁸Locke 1690, Chapter XXVII, 15.

identity. Apart from the fact that this presupposition might face counterexamples, the reading by itself gives no justification (or no positive reason) for accepting it. It is certainly not a basic piece of knowledge that accountability requires personal identity.²⁸⁹ It is thus far from obvious how we could move from the insight that norms of accountability track a certain relation to the further claim that what is being tracked is personal identity—not without already knowing the correct criterion of personal identity.

Either the *Prince and the Cobbler* is merely an illustration of a theory and not intended to convince us of what is being illustrated; *or* it is supposed to convince us of a theory, but provides insufficient reason for what it tries to establish.

4.2 Williams' Cases

Williams describes two further scenarios which, like Locke's *Prince and the Cobbler* and S. Shoemaker's *Brownson*, involve the separation of bodily and psychological continuity. I paraphrase them as follows:²⁹⁰

Case 1: A machine transfers person *A*'s psychology to the body of person *B* and vice versa.

Case 2: I will be tortured tomorrow. Before the torture, a doctor will cause me to forget the announcement of torture and everything I can now remember. I will even receive a different set of memories, belonging to another person now living. After the torture, I regain all my current memories, which during the torture are stored in a different brain.

²⁸⁹As I will mention later, Locke disagrees and takes the link between personhood and forensic unity to be analytic.

²⁹⁰Williams 1970, pp. 163, 167-8.

Williams mentions a number of reasons for why *Case 1* is plausibly described as an instance of body-switching where *A* and *B* go with their respective psychologies, i.e. the *B*-Body-person is *A*, and the *A*-body-person is *B*.²⁹¹ For example, if before the experiment, *A* decides that after the experiment, the *B*-body-person shall receive \$100,000, the *B*-body-person will, upon receiving the money, remember and approve of what she regards to be *her* choice.

Case 2 strikes Williams as terrible despite the psychological discontinuity:

“Fear, surely, would still be the proper reaction: and not because one did not know what was going to happen, but because in one vital respect at least one did know what was going to happen—torture, which one can indeed expect to happen to oneself, and to be preceded by certain mental derangements as well.”²⁹²

Case 2 thus supports the view that I go (or rather stay) with my body.

This, Williams notes,²⁹³ is puzzling. He thinks that if *A* in *Case 1* is to make a selfish choice, she will assign the money to the *B*-body-person who has her psychology. In *Case 2*, psychological discontinuity does nothing to alleviate my prospect of torture, but makes things worse:

“the predictions of the man in charge provide a double ground of horror: at the prospect of torture, and at the prospect of the change in character and in impressions of the past that will precede it.”²⁹⁴

Case 1 suggests a psychological criterion, while *Case 2* suggests a bodily criterion. But of course, *Case 2* is just one side of *Case 1*. Our intuitions are at odds.

²⁹¹Williams 1970, pp. 163-7.

²⁹²*Ibid.*, p. 168.

²⁹³*Ibid.*, pp. 168-9.

²⁹⁴*Ibid.*, p. 171.

One might think that this seeming defect in our intuitions has a harmless explanation. Maybe the cases actually differ in important respects. Williams mentions two options. First, only *Case 2* is presented in first-personal terms, while *Case 1* is third-personal. Maybe first-personal presentation causes fear when rational reflection would reveal that there is no reason to be afraid. But Williams maintains that not the pronouns, but acceptance of a principle like the following causes our reaction:

“my undergoing physical pain in the future is not excluded by any psychological state I may be in at the time [...]. In particular, what impressions I have about the past will not have any effect on whether I undergo pain or not.”²⁹⁵

Secondly, only *Case 1* mentions another individual to whom the relevant psychology is transferred. Williams thinks this cannot matter either. He considers a sequence from *A* becoming amnesiac to *A* switching psychologies with another individual *B*:

“(i) *A* is subjected to an operation which produces total amnesia;
(ii) amnesia is produced in *A*, and other interference leads to certain changes in his character;
(iii) changes in his character are produced, and at the same time certain illusory ‘memory’ beliefs are induced in him; these are of a quite fictitious kind and do not fit the life of any actual person;
(iv) the same as (iii), except that both the character traits and the ‘memory’ impressions are designed to be appropriate to another actual person, *B*;
(v) the same as (iv), except that the result is produced by putting the information into *A* from the brain of *B*, by a method which leaves *B* the same as he was before;
(vi) the same happens to *A* as in (v), but *B* is not left the same, since a similar operation is conducted in the reverse direction.”²⁹⁶

Williams argues that in (i), the *A*-body-person clearly remains *A*. It is *A* who becomes and remains amnesiac. Moreover, nowhere in the sequence

²⁹⁵[Ibid.](#), p. 169.

²⁹⁶[Ibid.](#), p. 172.

can a line be drawn such that the *A*-body-person ceases to be *A* —not even at the end of the sequence, where *A* switches psychologies with *B*. *A* is justified in fearing (i), and this fear can extend to (vi). The presence and explicit mention of *B* (*v*) does not alter our judgement that the *A*-body-person remains *A* throughout. If so, why should it matter that in (vi), *B* (or rather: the *B*-body-person) receives *A*'s psychology?

Williams considers two strategies to resist the diagnosis that *A* persists from (i) to (iv). First, conventions could settle how long *A* exists in the sequence. Specifically, he considers the view that *A* survives as the best candidate for being *A*. If so, (*v*) and (vi) need not be treated on par. In (*v*), *A* is the *A*-body-person because this is the best *A*-candidate, while in (vi), an even better candidate is around: *A* is the *B*-body-person. Second, we might maintain that *A* indeed exists in (i) and has ceased to exist in (vi), but that one or more steps in the sequence are borderline cases where it is indeterminate whether *A* still exists.

Williams dismisses the conventionalist solution. Any conventionalist person boundary between (i) and (vi) will fail to delimit *A*'s fear. For example, if through convention we determine that the presence of someone else in (iv) makes all the difference, Williams thinks it is implausible that *A*'s fear

“can be rationally diverted from the fate of the exactly similar person in (vi) by his being told that someone would have a reason in the latter situation which he would not have in the former for deciding to call another person *A*.”²⁹⁷

In other words, a convention according to which at one point in the sequence, *A* will cease to be the *A*-body-person will not prevent *A*'s fear

²⁹⁷Williams 1970, p. 178.

from extending beyond the conventionalist person boundary.

The indeterminacy line leads to a related complication. A situation where it is conceptually indeterminate whether I survive has “no comprehensible representation in my expectations and the emotions that go with them.”²⁹⁸ If the situation involves me, I can engage in *projective imagination* and determine what the experience would be like for me. But if it is indeterminate whether the situation involves me,

“fear [...] seems neither appropriate, nor inappropriate, nor appropriately equivocal. Relatedly, the subject has an incurable difficulty about how he may think about [situation] S. If he engages in projective imaginative thinking (about how it will be for him), he implicitly answers the necessarily unanswerable question; if he thinks that he cannot engage in such thinking, it looks very much as if he also answers it, though in the opposite direction. Perhaps he must refrain from such thinking; but is he just refraining from it, if it is incurably undecidable whether he can or cannot engage in it?”²⁹⁹

In other words, it seems incredibly puzzling which attitude I should take towards borderline cases. If I fear them, I regard the experiencer as determinately me. If I don't, I regard her as determinately not me. Both attitudes conflict with the supposition of indeterminacy.

I will come back to these responses. For the moment, I note that Williams concludes that neither the first-personal pronoun in *Case 2* nor the mentioning of another individual in *Case 1* explain why our judgments seem in conflict.

“Thus, to sum up, it looks as though there are two presentations of the imagined experiment and the choice associated with it, each of which carries conviction, and which lead to contrary conclusions”³⁰⁰.

²⁹⁸[Ibid.](#), p. 175.

²⁹⁹[Ibid.](#), pp. 176-7.

³⁰⁰[Ibid.](#), p. 178.

Williams closes with some skepticism about our reactions to *Case 1*.

While they neatly support a psychological criterion,

“this neatness is basically artificial; it is the product of the will of the experimenter to produce a situation which would naturally elicit, with minimum hesitation, that description.”³⁰¹

In close and equally possible scenarios, such as (*v*), or a situation where two persons with *A*'s psychology are produced, our reactions would be much different. We should thus take our judgements about *Case 1* with a grain of salt.

4.3 Transplant Intuitions Are Off-Track

I now formulate some worries about Williams' evidence on the individual *Cases*, before articulating broader reservations about transplant intuitions.

4.3.1 Quality Of Evidence: *Case 1*

On *Case 1*, Williams comments:

“the *B*-body-person will not only complain of the unpleasant treatment as such, but will complain (since he has *A*'s memories) that that was not the outcome he chose, since he chose that the *B*-body-person should be well treated; and since *A* made his choice in selfish spirit, he may add that he precisely chose in that way because he did not want the unpleasant things to happen to *him*. The *A*-body-person meanwhile will express satisfaction both at the receipt of the \$100,000, and also at the fact that the experimenter has chosen to act in the way that he, *B*, so wisely chose. These facts make a strong case for saying that the experiment has brought it about that *B* did in the outcome get what he wanted and *A* did not. It is therefore a strong case for saying

³⁰¹Williams 1970, p. 179.

that the *B*-body-person really is *A*, and the *A*-body-person really is *B*; and therefore for saying that the process of the experiment really is that of changing bodies. For the same reason it would seem that *A* and *B* in our example really did choose wisely [...]. This seems to show that to care about what happens to me in the future is not necessarily to care about what happens to this body (the one I now have); and this in turn might be taken to show that in some sense of Descartes's obscure phrase, I and my body are 'really distinct'³⁰².

These comments, I argue, should strike us as puzzling. The *A*-body-person will surely approve of the experimenter's compliance with *someone's* earlier choice to allocate the money to the *A*-body-person, who surely gets what she wants now. And the *B*-body-person will certainly complain about the experimenter's failure to comply with *someone's* earlier choice to allocate the money to the *B*-body-person. But does the *A*-body-person get something *she* chose earlier? And is the *B*-body-person deprived of something *she* chose earlier?

If the *A*-body-person claims that it was *her* choice that the money shall be allocated to the *A*-body-person, she would know something which the thought experiment was supposed to establish: whether persons go with their psychology, and decision-maker *B* is now the *A*-body-person. The *A*-body-person would assume or presuppose that persons go with their psychology. But by itself, this would not provide reasons to think that this view is correct. Thus, it is mysterious how the fact that *A* but not *B* got what she wanted gives a "strong case" that the *B*-body-person is *A*, and the *A*-body-person is *B*. The quoted considerations are *circular* in presupposing what was supposed to be shown.³⁰³

³⁰²*Ibid.*, p. 164.

³⁰³Analogous remarks apply to Williams' variants of *Case 1*. For example, suppose *A* chooses money for the *A*-body-person, and *B* conversely. If the *A*-body-person

Now, Williams mentions that *A* and *B* shall choose in a *selfish* manner. Presumably, good selfish decisions maximize *one's own* utility over time. It is noted that both *A* and *B* “chose wisely” (but only *B* had her choice implemented). This suggests that the decision-makers really succeeded in making a selfish choice, one that is good *for them* respectively. Since their choices favoured other bodies, it seems that caring selfishly for *oneself* is “not necessarily to care about [one's] body”. Williams thinks these points support his identity judgements. I am convinced that we should have serious questions about the informativeness of the claim that *A*'s and *B*'s choices are both selfish and wise.

One reason for skepticism is described by Noonan. He argues that the object of selfish concern will depend on the individuals' views on personal identity.³⁰⁴ The crucial step in Williams' discussion is to move from the *A*-body-person's memory of making the choice, to declaring it a choice “he, *B*,”³⁰⁵ made. Noonan distinguishes the choice the *A*-body-person *recalls* having expressed from the choice the *A*-body-person *actually* made. The *A*-body-person might seem to recall having made *B*'s choice. But she will regard *B*'s choice as a choice she, the *A*-body-
receives the money,

“[i]t looks, then, as though the *A*-body-person had gotten what he wanted, but not what he chose, while the *B*-body-person has gotten what he chose, but not what he wanted. So once more, it looks as though they are, respectively, *B* and *A*” (Williams 1970, p. 165).

But the question arises on what grounds the *A*-body-person should be entitled to say that this distribution is not what *she* chose, and vice versa for the *B*-body-person? Granted, the *A*-body-person has *B*'s psychology, and thus memories of choosing differently. But for whose choice this was, these memories only matter if a psychological account is presupposed. And whether a psychological account is true is precisely what we wanted to find out.

³⁰⁴Noonan 1989, section 10.3.

³⁰⁵Williams 1970, p. 164.

person, *actually* made for herself only if she, the *A*-body-person, believes in a psychological criterion of personal identity.

In contrast, suppose the *B*-body-person believes in a bodily rather than a psychological criterion. Then she will regard it unfortunate that the experimenter did not bring about the outcome she has a memory of choosing. But importantly, she will regard this memory as *illusory*. It was not her, the *B*-body-person, who made the choice she is recalling now. Instead, it was (in her view) someone else: *A*.

This shows that who benefits from *A* and *B*'s *selfish* choices depends on who they think they will be. And which earlier choices the *A*- and *B*-body-persons regard as their own depends on who they think they were. For example, if *B* believes in a physical criterion, and the *A*-body-person finds herself with *B*'s psychology, then upon receiving the money, she will be delighted, yet report, when confronted with what happened, that her memory of making this choice is only apparent. In line with her belief in the physical criterion, she will regard herself as identical to *A*. She will say that she cannot remember her *actual* choice, but that the choice she *seems* to remember benefits her now.³⁰⁶

According to an unpublished response by S. Shoemaker, Noonan's criticism falsely assumes that the individuals' intuitions are a function of their theory of personal identity. Whiting reports that for S. Shoemaker, this

“completely misunderstands the method of cases’, which works by trying to get one’s opponent first to see that her intuitions

³⁰⁶Maybe regardless of what the *A*- and *B*-body-person's believe about personal identity, they have a *tendency* to judge on the basis of their apparent memories, and hence in line with the psychological criterion. But if they subscribe to a bodily criterion, Noonan says, they will suppress this tendency.

about personal identity in particular cases conflict with her theory about personal identity and then to modify either her theory or her judgments about particular cases accordingly.”³⁰⁷

Within the S. Shoemaker picture, intuitions and theory can conflict. It is up to the reasoning individual to resolve this conflict; she need not stick to her theory.³⁰⁸ For example, even if *A* and the *B*-body-person who inherits *A*'s psychology believe in a physical criterion, it is not the case that the *B*-body-person necessarily suppresses her belief that *her* choice to give *A* the money was respected. In light of the case, she might well revise her picture of personal identity in favour of a psychological criterion. Thus, if S. Shoemaker is right, the individuals' responses are informative, because they are not just functions of antecedent beliefs. They are reflective judgements on how to harmonize antecedent theory and occurrent intuitions.

Noonan receives support by Whiting, who argues convincingly that S. Shoemaker's point does not help. Firstly, when intuitions and theory *conflict*, Noonan's point stands that the individual can consistently dismiss intuitions that conflict with her theory. The individual's judgement then is contingent on her theory. Secondly, anytime intuitions *cohere* with the individual's theory, this might be so *because* they are informed by the theory. The intuition then is merely a symptom of the individual's endorsement of the theory, but goes no way to *support* it. As a result, the judgements on *Case 1* are not neutral. We cannot rule out—in fact it is likely—that they are merely expressive of an antecedent view of personal identity. The problem is that

³⁰⁷Whiting 1999, fn. 12.

³⁰⁸Beck 1998, p. 215 makes a similar complaint against Noonan.

“this tells us primarily what the subject *thinks* or *feels* about the nature of personal identity and only secondarily (if at all) about the nature of personal identity *itself*.”³⁰⁹

Indeed, Noonan concludes that the reports of the *A*- and *B*-body-persons do nothing to support a psychological criterion:

“in so far as the *A*-body-person and the *B*-body-person are likely [...] to make statements that appear to provide support for a psychological criterion of personal identity, this will be so only if they are themselves believers in such a criterion; any criterion of personal identity, no matter how bizarre, could be given a similar appearance of support by imagining what would be said by people crazy enough to accept it.”³¹⁰

He concedes that it *might* be correct to describe *Case 1* as body-switching. The point is that whether the individuals in the Case possess attitudes that could support a psychological criterion depends on what they themselves believe about the topic. When the *B*-body-person approves of *A*'s earlier choice, she is presupposing a psychological criterion. Williams' discussion does not establish that we *should* make this presupposition.

I conclude that William's remarks on who is who are circular in presupposing a criterion rather than providing evidence for one. Moreover, Noonan's criticism brings out that without further discussion, these presuppositions are arbitrary and unwarranted.

4.3.2 Quality Of Evidence: *Case 2*

Williams is not maximally explicit about one difference between the *Cases*. *Case 2* lacks an analogue to the distinction between *A* and the

³⁰⁹Whiting 1999, p. 446.

³¹⁰Noonan 1989, p. 179.

A-body-person in *Case 1*. While the first draws a terminological distinction between the pre-experimental, decision-making subjects *A* and *B*, and the post-experimental individuals *A*-body-person and *B*-body-person, the second case is described in terms of ‘I’ throughout. Not only is the second case first-personal, it also uses the same pronoun *across the whole case*. As mentioned, Williams thinks this does not matter; regardless of the pronoun used in the presentation, we see that the pain suffered through amnesia will be ours.

Against this, usage of *the same* pronoun throughout threatens to be just as circular as William’s remarks on *Case 1*. Whether it will be *me* who undergoes pain is precisely what is in dispute.³¹¹ It is thus unclear why we should be entitled to think that I persist throughout the loss of my psychology, and that we should use the same pronoun accordingly. *Case 2* is described in a question-begging way.

The principle referred to by Williams—that future pain is not excluded by any psychological state at that time—does not establish that I persist throughout *Case 2*. Whiting puts the point as follows:³¹² psychological continuity theorists can simply accept the principle; but they are free to deny that the psychologically discontinuous pain-sufferer will be *me*. This is compatible with the principle that if a future individual *is* me, I can fear his pain regardless of his psychology.

This has implications for Williams’ sequence argument. First, we should already have reservations about the starting point that *A* survives in the first step, (*i*). It plainly begs the question against psychological

³¹¹Beck 2014, p. 191.

³¹²Whiting 1999, p. 449.

continuity theory.³¹³ Second, for psychological continuity theorists, the sequence is underdescribed. They can maintain that whether a line can be drawn beyond which *A* ceases to exist depends on the exact psychological changes under consideration. If *(i)* features only the loss of memories of past experiences, then it is indeed plausible that *A* survives. If the changes are more drastic and feature no psychological continuity whatsoever, then we could deny that *A* survives even *(ii)* or *(iii)*.³¹⁴

I conclude that William's discussion of *Case 2* too begs the question, this time in favour of a bodily criterion. To sum up, I think his remarks suffer from the same shortcomings as the *Prince and the Cobbler*. Either they are *illustrations* of a criterion of personal identity and not intended to convince us of one.³¹⁵ What I have deemed bad evidence then would not be intended as such, but merely part of the illustration. Or the remarks *are* intended to convince us of a criterion, but do so on insufficient and circular grounds.

4.3.3 Quasi-Attitudes

In his discussion of borderline cases, Williams mentions something that might illuminate his comments on the *Cases*. He concedes that, unless one is exceptionally egoistic, being concerned about an event does not presuppose that oneself figures in it.

³¹³*Ibid.*, p. 452.

³¹⁴Garrett 1998, pp. 54-5. For similar points, see Oyowe 2013, pp. 264-5, Herdova 2016, pp. 327-330, 335-6. Noonan 1989, pp. 185-6 argues for drawing the line no later than between *(iv)* and *(v)*.

³¹⁵Williams then would have *illustrated* two different criteria—no wonder the pictures diverge.

“There are some emotions, however, which I will feel only if I will be involved in [situation] *S*, and fear is an obvious example”³¹⁶.

This suggests that the reason why Williams thinks that the individuals’ reports in the *Cases* are informative about personal identity is the following. In *Case 1*, *A* fears that pain could be inflicted upon the *B*-body-person. And *fear* can be felt by *A* only if she believes that *she* will be affected. The same goes for the individual in *Case 2* who fears the pain of the amnesiac getting tortured. Moreover, in *Case 1* the *B*-body-person, upon receiving the money, will feel *compensated* for *A*’s wise choice. Maybe this feeling of compensation can only be had for one’s own investments and achievements. Thus, the *B*-body-person must regard *A* as her. More generally, the individuals report attitudes that are essentially egocentric, and this might help us to draw inferences about personal identity.

Unfortunately, this does not make Williams’ comments on the *Cases* more plausible. For one thing, Noonan and Whiting’s point stands: it depends on the subject’s theory whom she regards as herself. Structuring egocentric concerns around some past and future selves rather than others indicates what the subject’s theory is, but does not *justify* it. Moreover, there is a further problem for taking these attitudes as evidence for personal identity. There could be attitudes that are *exactly* like those Williams regards as essentially egocentric, except for that it is not conceptually required that they are only structured around oneself. Their possessors can structure them around others as well.

³¹⁶Williams 1970, p. 175.

These attitudes would be analogues to S. Shoemaker’s notion of quasi-memory, i.e. first-personal knowledge of a past experience where the knowing subject and the earlier, experiencing subject need not be identical.³¹⁷ S. Shoemaker argues that in resembling memory, quasi-memories are causally connected—in his terminology: M-connected, thereby establishing quasi_c-memories—to the experience that is being quasi-remembered. M-connections

“create a presumption that he, that same person, experienced the event or did the action, and therefore a presumption that the quasi_c-memory was actually a memory.”³¹⁸

The presumption is suspended if the quasi_c-memory chain branches. S. Shoemaker defends non-branching quasi_c-memory as the criterion for personal identity. If x and y are related in this way, they are related

“in essentially the same ways as people’s memories are generally connected with their own past experiences and actions”³¹⁹,

which entitles us to conclude that x is the same person as y , without falling prey to Butler’s circularity objection (1.2.2).

Now, I am here merely availing myself of quasi-attitudes. I am not endorsing the claim that non-branching quasi-memory constitutes personal identity. My focus is not on assessing psychological continuity theory or S. Shoemaker’s version thereof. I am concerned with one important consideration in this assessment: transplant intuitions.³²⁰ And for them, I am convinced that the availability of quasi-attitudes raises problems.

³¹⁷S. Shoemaker 1970, p. 271.

³¹⁸*Ibid.*, p. 279.

³¹⁹*Ibid.*, p. 282.

³²⁰One quick reminder: as I use ‘transplant intuitions’, they need not refer a literal transplant of an organ (4.1). It suffices that psychological and bodily continuity become separated, for example as in the *Prince and the Cobbler* and the *Cases*.

Earlier, I suggested that anticipation has a quasi-counterpart (2.5.2). There is a sense in which I can regard a future self as *me*, as the object of reflexive thought, while staying neutral on whether ‘we’ are numerically identical.³²¹ I also explained Brink’s notion of *quasi-prudence*, which requires benefactor and beneficiary to be psychologically continuous but not necessarily identical (3.1.2).

The implications for Williams are the following. Since these quasi-notions are perfectly like their genuine counterparts—except for not being necessarily about oneself—they are also phenomenologically alike. But if the egocentric and the quasi-notions are qualitatively indistinguishable, then the *B*-body-person’s seemings of compensation for *A*’s choice, or my reported fear of the torture, show nothing. Of course, we could just stipulate that the *Cases* involve genuinely egocentric attitudes. But then the cases would not reveal anything new to us. Without the stipulation of egocentricity, we are never entitled to believe that the individual in question has the egocentric rather than the quasi-attitude. Seemings of anticipation, fear, compensation can never actually license an inference about numerical identity, because for all we know it could be a quasi-attitude that is entirely neutral on this matter.

Proponents of either a physical or a bodily criterion can find this plausible. Here are two examples. First, consider Parfit’s *Teletransportation*. For an animalist, it is a case of non-identity, given that the biological organism created on Mars is distinct from the organism destroyed on Earth. But it is assumed that the operation secures and transfers the Earthling’s psychology perfectly well. It should be easy for the Earthling

³²¹Velleman 1996, especially sections 4-5.

to projectively imagine the Martian's experiences. The animalist will be perfectly fine with allowing projective imagination to transcend person boundaries.

Secondly, psychological continuity theorists accept that psychological connectedness comes in degrees. There is personal identity if psychological connections are sufficiently strong, and personal distinctness if the connections are sufficiently weak. But in between, the degree of connectedness neither suffices for identity nor is low enough for distinctness. Whiting argues that it is not implausible that a subject quasi-projectively imagines being in such a state, and comes to quasi-anticipate experiencing it.³²² Here, the theory need not deny that this imagination is egocentric: it can understand an egocentric or selfish attitude as being structured around psychological continuers rather than identical future selves. Moreover, in such a case the subject need not share Williams' worries about a lack of "comprehensible representation in her expectations"³²³ and an "incurable difficulty about how [s]he may think about [the situation]"³²⁴. Rather than feeling ambivalent concern for the future self that comes close to regarding her as distinct and thus unsuitable for egocentric attitudes, she could just *attenuate* her attitudes in accordance with the degree of psychological connectedness she expects.

I conclude that setting aside the circularity worries, once we entertain the possibility that the individuals in the *Cases* have quasi-attitudes, their responses' evidential value is diminished.

³²²Whiting 1999, pp. 456-7.

³²³Williams 1970, p. 175.

³²⁴*Ibid.*, p. 177.

4.3.4 Divergence

A further worry arises if we consider the judgements from the *Cases* taken together. Suppose we *can* take the individuals' attitudes as evidence in the way Williams suggests. And suppose Williams is right that our intuitions are not sidetracked by presentational features. The resulting overall picture is worrisome. Our judgements diverge: *Case 1* supports a psychological criterion, while *Case 2* suggests a physical criterion. And it is mysterious which feature of the *Cases* could account for the divergence. This "conundrum"³²⁵ suggests that our intuitions are a mess.

This threatens to render the position that transplant intuitions are evidence for personal identity dubious. Williams believes both that transplant intuitions are evidence *and* that they result in divergent judgements. But one might think that the conjuncts of this compound proposition undermine each other. If transplant intuitions are evidence for personal identity, then they would lead to well-behaved judgements that do not diverge and vary inexplicably. Moreover, if transplant intuitions lead to divergent, contradictory judgements about the criterion for personal identity, then they seem to provide poor evidence on this matter.

This shows that acceptance of the compound proposition—transplant intuitions are evidence for personal identity *and* quasi-prudential concern gives rise to divergent judgements on the topic—amounts to an unstable position. If one conjunct holds, the other cannot hold. One conjunct needs to go.³²⁶

Since it is unclear how the diverging judgements could be explained

³²⁵Johnston 1987, p. 65.

³²⁶In 4.6, I discuss Sider's and Eklund's positions where this wouldn't follow.

away, I argue that we should see the diagnosis of divergence as a *reductio*. Suppose that transplant intuitions provide evidence on personal identity. As Williams' discussion demonstrates, this leads to divergent intuitions, supporting a psychological criterion in one case but a physical one in another. Again, we grant Williams that in neither case are intuitions sidetracked by irrelevant features such as presentation. In other words, the proposed evidential basis leads to inexplicable, apparently unmotivated divergence in our judgements. The starting assumption must be false. Transplant intuitions do not provide evidence on personal identity.

I will now examine the implications of these results for the justifiedness of beliefs that are based on transplant intuitions.

4.4 Debunking Transplant Intuitions

I am convinced that the foregoing sections motivate a debunking argument against transplant intuitions. Debunking arguments provide undermining defeaters.³²⁷ Suppose I believe p on the basis of being in state M , e.g., having a belief, percept, or intuition. A *defeater* is a state M^* that, together with M , does *not* give me reason to believe p .³²⁸ A defeater is *undermining* if the reason why M does not give me reason to believe p is that M^* suggests that the connection between M and belief in p is faulty: p could be false and we would still be in state M .

Debunking arguments have the following components: a *causal* premise according to which a belief-forming method causes a subject to believe p , an *epistemic* premise according to which the method is not truth-

³²⁷Kahane 2011, pp. 104-6.

³²⁸For these definitions, see Pollock and Cruz 1986, pp. 195-7.

tracking, and the conclusion that the subject is not justified in believing p on the basis of the method. p could be false, but the subject would still believe it. The undermining defeater here is the second-order belief that belief in p is the result of a non-truth-tracking belief-forming process.

I am suggesting the following debunking argument (illustrated on the *Prince and the Cobbler*):

Causal premise: Our belief that the prince goes where his soul/body goes is explained by transplant intuitions.

Epistemic premise: Transplant intuitions are off-track, i.e. do not reliably track truth.
Therefore,

Conclusion: Our belief that the prince goes where his soul/body goes is unjustified.

Note that this argument applies not just to those transplant intuitions that seem to support psychological continuity theory, such as in the *Prince and the Cobbler* and William's *Case 1*. It equally applies to endorsements of bodily criteria on the basis of transplant intuitions, such as in *Case 2*.

With the causal premise, I am not assuming that transplant intuitions are the *only* possible reason for believing, e.g., that the prince goes where his soul goes. I am merely assuming that in the context of the case, it is these intuitions that cause the belief. In the *Prince and the Cobbler*, the judgement partly rests on beliefs about who is accountable for what.

Reflections on Locke and Williams revealed that there is something wrong with transplant intuitions. They fail to provide non-circular reasons for the respective belief. They also support different criteria in different cases, without there being features in the cases that could plausibly explain this divergence. They take the subject's future- and past-

directed attitudes, which feed into her transplant intuitions, as evidence for personal identity. Once we realize that these attitudes could be of the quasi-variety and extend to distinct persons without differing phenomenologically, it becomes apparent that they cannot play this evidential role. We could have these attitudes in the absence of personal identity. A subject's report of a distinctive set of attitudes towards a past or future subject thus does not establish that these subjects are numerically identical. The transplant intuitions that are shaped by these attitudes are not tracking truths about personal identity.

We should thus conclude that our endorsement of a theory of personal identity on the basis of transplant intuitions is unjustified. Note that this result is compatible with justified endorsements on other grounds – considerations that do not depend on transplant intuitions. Moreover, the result is compatible with transplant intuitions introducing positions into a debate, and thereby setting the stage for further argument and post-hoc rationalization.³²⁹ That is, these intuitions can *generate* hypotheses we go on to debate in light of further reasons. But they go no way to *justify* these hypotheses.

Transplant intuitions are unhelpful for taking a stance in the debate between psychological and physical criteria. This does not mean that either side has been refuted, only that neither can be supported by means of transplant intuitions. As a consequence, congruence with transplant intuitions is not as such a point *in favour* of a theory of personal identity. And if a theory of personal identity seems in tension with transplant intuitions, this is not as such a reason to *discredit* the theory. Considerations

³²⁹Kahane 2011, p. 106.

of this kind are neutral with regards to justification.

4.5 Debunking And Objectivism About Personal Identity

I have to discuss a seeming presupposition of the foregoing. Olson draws a distinction between personal identity and *being the same person*. Personal identity is the numerical identity of persons over time. *Being the same person* refers to unity for moral or practical purposes. One example for spelling out the latter is:

“*x* is at time *t* the same person as *y* is at a later time *t** if and only if *x* ought to be prudentially concerned, at *t*, for *y*’s well-being at *t**; and *y* is responsible, at *t**, for what *x* does at *t*; and it is natural and right at *t** to treat *y* as if she were *x*.”³³⁰

Olson points out that *being the same person* need not coincide with personal identity. For one thing, fission shows that the relations differ in formal properties:³³¹ *being the same person* seems intransitive because each fission product could stand in this relation to its predecessor. Moreover, the relation captures a social role an individual occupies. One and the same role can be occupied by distinct individuals, e.g., replicas could be related by *being the same person* but not personal identity. In terms of Williams’ *Case 1*, it might well be that the *B*-body person *is the same person* as *A*. But it is an open question whether they are numerically identical. *A*’s social role before the operation could be played by a numerically distinct person afterwards. And *Case 2* could be seen as an

³³⁰Olson 1997, p. 66.

³³¹*Ibid.*, pp. 68-9.

example where someone remains numerically identical but fails to *be the same person* as the tortured individual.³³²

Olson's discussion has at least two important aspects: first, the distinction between personal identity and sameness for practical purposes; second, the claim that the former obtains *independently* of the latter. We can appreciate Olson's distinction without endorsing the independence claim. In fact, Eklund draws a very similar distinction (with different terminology) and goes on to call it the "naïve (in the sense of simple and intuitive) view"³³³ that the two distinguished topics are intimately connected: personal identity is determined by the nature of the entities at the centre of the practical concerns figuring in *being the same person*, and vice versa.

One direction of this "naïve" dependence relation can be expressed through the *moral relevance assumption* about persons:³³⁴ it is part of our conception of personhood that those entities around which we should structure our prudential and moral concerns are the persons. Dummett writes:

"We can easily imagine people who use different criteria from ours [...]. Precisely what would make the criteria they used criteria for *personal identity* would lie in their attaching the same consequences, in regard to responsibility, motivation, etc., to their statements of personal identity as we do to ours."³³⁵

As Eklund explains, if one accepts the moral relevance assumption, *being the same person* wouldn't be independent and irrelevant, but of crucial

³³²Olson acknowledges that animalism might make personal identity seem less important for practical purposes; e.g., *ibid.*, ch. 3, sect. VIII; 2013, p. 82; *forthcoming*, section 6.

³³³Eklund 2004, p. 489.

³³⁴*Ibid.*, p. 503.

³³⁵Dummett 1973, p. 358.

importance for giving an account of personal identity.³³⁶ And it would be a competitive disadvantage of Olson’s view if the concerns expressed through transplant intuitions track psychologies rather than human organisms.³³⁷

Now, if we deny Olson’s claim that personal identity is independent from the practical concerns mentioned in *being the same person*,³³⁸ then it is not obvious anymore that transplant intuitions are bad truth-trackers. For example, Locke takes it to be analytic that personhood is a forensic concept which appropriates actions and their merits.³³⁹ The prince is the same person as the cobbler’s body with the prince’s soul if we can hold the latter responsible for actions of the former. If this conditional holds, surely the intuition that the cobbler with the prince’s consciousness would indeed be responsible for the prince’s actions is high-quality evidence for the claim that the prince is the cobbler, and thus that persons go where their consciousness goes.

Whiting suggests that denying the independence thesis would also significantly improve the informativeness³⁴⁰ of Williams’ cases. If

³³⁶Eklund 2004, pp. 504-5. See, e.g., Baker: “it is deeply wrong to divorce identity from moral responsibility, or from what we care about” (2008, p. 13).

³³⁷In defense of Olson, I respond that his view concerns the *ontological* question of which person candidates exist, not the *semantic* question of what ‘person’ is true of. It might be that moral relevance is built into the semantics of ‘person’. However, in this picture Olson’s project of settling which potential referents for ‘person’ there are would come under pressure only if moral considerations (to which psychological continuity might seem more relevant than sameness of organism) affects what exists.

³³⁸For example, D. Shoemaker suggests that *being the same person* could make essential reference to personal identity. The plausibility of a theory of the latter could then “increase or decrease depending on the extent to which it fits with our practical concerns” (2016, p. 316).

³³⁹*Pace* Eklund 2004, pp. 495-6, I read Locke as an instance of, not a counterexample to, the “naïve” view.

³⁴⁰However, Whiting calls it “William’s dilemma” (1999, p. 460) that the independence thesis is essential to other parts of his discussion: when dismissing the idea that there are borderline cases between (i) and (vi), he seems to suppose that it is a

“part of what *makes* some future self *his* future self is the fact that he *now* has certain attitudes toward it, then his attitudes toward that future self would not be *mere evidence* that he *regards* that self as his own: these attitudes might, in suitable circumstances, be *partly constitutive* of its *being* his future self. In that case, these attitudes could function directly as evidence about which future selves are *in fact* his own.”³⁴¹

A similar point is often made with regards to evolutionary debunking arguments in metaethics. According to these arguments, we hold our current moral beliefs partly because they proved to be adaptive in the course of evolution. However, evolutionary selection proceeds irrespective of moral truth. Our moral beliefs thus seem unjustified. Kahane argues that this worry is pressing only for *objectivism* about moral values:

“anti-objectivist views claim that our ultimate evaluative concerns are the source of values; they are not themselves answerable to any independent evaluative facts. But if there is no attitude-independent truth for attitudes to track, how could it make sense to worry whether these attitudes have their distal origins in a truth-tracking process?”³⁴²

If the moral facts depend on us, why should we entertain the epistemic premise that evolution has selected moral beliefs regardless of their truth? For our purposes, the analogous point raised by Whiting is: if personal identity depends on us, why should we suppose that the attitudes captured in transplant intuitions give rise to beliefs regardless of their truth? Presumably, attitudes of concern, compensation, and moral responsibility would be good guides to personal identity.

mistake to have (quasi-)egocentric concern for one’s psychological continuer. Whiting thinks Williams must hold that such attitudes do not fit independent facts. To make this claim, Williams needs the independence thesis.

³⁴¹*Ibid.*, p. 446.

³⁴²Kahane 2011, p. 112.

If debunking arguments would presuppose objectivism about the target domain, then my debunking argument could only be launched against transplant intuitions when employed to support objectivist, i.e. response-independent accounts of personal identity. While there are many such positions,³⁴³ this would still narrow down the scope of the argument.³⁴⁴

I deny that endorsing anti-objectivism suffices to escape the respective debunking argument. The authors seem to think that

Epistemic premise. X is an off-track process.

is not true if what is being tracked depends on the individual(s) doing the tracking. But of course, even if anti-objectivism is presupposed, worries about truth-tracking can arise. If I believe p based on flipping a coin, my belief can be debunked regardless of whether or not objectivism about p -matters is presupposed.³⁴⁵

If anti-objectivism about the moral domain is presupposed, truth-tracking worries about adaptive beliefs can arise. Evolution gives rise to *all kinds of* attitudes.³⁴⁶ Presumably, even if moral matters depend on us, evolution will have selected some attitudes that accord with anti-objectivist moral facts, and others that conflict with them. It is thus a fair question whether a particular attitude we possess partly due to

³⁴³Objectivism seems to be endorsed, e.g., in S. Shoemaker 1984, esp. p. 101, Olson 1997, 2007, Merricks 2001b, Snowdon 2014.

³⁴⁴For more or less explicit endorsements of anti-objectivism, cf. Nozick 1981, p. 69, Unger 1990, p. 66, Parfit 1995, p. 25, Sider 2001a, Braddon-Mitchell and Miller 2004, Miller 2013, Kovacs 2016.

³⁴⁵For reasons like this, Joyce (2016, pp. 145-6) now denies that debunking arguments presuppose objectivism.

³⁴⁶Cf. Street 2006, p. 153: “each of us starts out with a vast and complicated set of evaluative attitudes”. She is clear that judgements on anti-objectivist matters can be mistaken, e.g., if they are not made in reflective equilibrium, or would not “withstand scrutiny from the standpoint of our other evaluative judgments” (*ibid.*, p. 154).

selective pressures is in line with anti-objectivist norms and values. We cannot just take mere possession of the attitude as indicative of moral truths.

Similarly, even if we follow Whiting in taking personal identity to depend on concerns figuring in *being the same person*, truth-tracking worries about transplant intuitions can arise.³⁴⁷ Our person-related practices fail to determinately track either psychological or bodily continuity. D. Shoemaker points out that a whole range of concerns plausibly figure in *being the same person*: prudential concern, moral responsibility, social treatment, compensation, anticipation, emotional patterns, etc.³⁴⁸ He argues that they are much less unified than commonly suggested. In particular, for each of them, different continuities matter in different ways. In all cases, it's ownership of an experience, action, or trait that matters for the practice. But ownership comes apart from identity and depends on different kinds of continuities.

Social treatment might appear to presuppose psychological continuity.³⁴⁹ Some would surely treat the cobbler with the prince's soul as the prince. On the other hand, maybe the cobbler could get arrested for impersonating a prince! In the spirit of *Case 2*, D. Shoemaker highlights contexts in which social treatment tracks not psychological but biological continuity.³⁵⁰ For example, parental love for one's child is independent

³⁴⁷For example, Johnston thinks that personal identity is a response-dependent concept (1989; 1993), rejects projectivism, i.e., the view that judgements about personal identity straightforwardly determine facts about personal identity (1993, p. 108), and believes that subjects make projectivist errors when thinking about personal identity (1987, pp. 69-70).

³⁴⁸D. Shoemaker 2016, p. 317.

³⁴⁹For example Johnston 1987, p. 82.

³⁵⁰See also Snowdon 2014, pp. 134-5.

of its psychology. And when illness leads to psychological discontinuity, the ill individual retains her property and remains in family relations.

Other examples relate to moral responsibility and compensation. It is a platitude that I can only be responsible for my own actions, and be compensated only for a burden I myself carried. It might thus appear that in order to be responsible for an action, I must be identical with its agent, and in order to be compensated, I must be identical to the burden-carrier, presumably by virtue of being psychologically connected to him. But D. Shoemaker denies this.³⁵¹ What matters is my *ownership* of the action or burden, and my duplicate could own my past actions or burdens without being identical to me. Although one might think that psychological continuity extends ownership of actions or burdens across time, it is not sufficient: severe psychological transformations, e.g., drifting into a fugue state, are compatible with psychological continuity if they unfold *gradually*, yet fall short of securing ownership of action. Neither is psychological continuity necessary: sudden psychological discontinuity, e.g., from a head-trauma, can leave ownership of action unaffected.

D. Shoemaker discusses further examples, and in conclusion discards the claim “our practical concerns constitute a unified set that is grounded on psychological continuity”³⁵² or some other relation in which personal identity might consist. When thinking about *being the same person*, “identity is the reddest of herrings.”³⁵³

According to D. Shoemaker, different attitudes require different continuity relations to be reasonable. But even worse, we might think that

³⁵¹D. Shoemaker 2016, pp. 318-9.

³⁵²*Ibid.*, p. 317.

³⁵³*Ibid.*, p. 323.

one and the same attitude requires different continuity relations, depending on its object. Whiting invites us to reflect on a variation of *Case 2* in which I, instead of getting tortured, get jilted by my lover.³⁵⁴ Suppose that while I dread pain in *Case 2*, I report indifference about my jilting lover since at the moment of jilting, all my memories and dispositions are gone.

First, Whiting shows how this report can be given two plausible but contradictory explanations. One is that I identify with my psychological continuer and do not care about whether my current lover remains with my body once my psychology is gone. Or I identify with my physical continuer while maintaining that whether something, such as the presence of my lover, is good for me depends on my psychological states at that time. By itself, the report does not clearly favour either explanation.

Second, the example demonstrates that whether concern goes with a physical or psychological criterion

“might differ if we were considering different sorts of pains and rewards. [...] [S]uch attitudes are not brute data but highly contingent phenomena sensitive to a wide range of individual and cultural factors”³⁵⁵.

Amongst others, the insight that plugging in different pains and rewards into one and the same case suggests different criteria of personal identity leaves us wondering which sorts of pains and rewards are the *right* indicators for *the* criterion of personal identity.

What matters for my purposes is that if we had antecedently thought that these practices inform us about personal identity, for example by

³⁵⁴Whiting 1999, p. 448.

³⁵⁵*Ibid.*

sympathizing with the anti-objectivist claim that personal identity depends on us, then the problem arises that these practices point in very different directions. Even if personal identity depends on us, it is a mistake to think that our transplant intuitions which reflect our attitudes and concerns—accountability in the *Prince and the Cobbler*, compensation, self-referential concern, and anticipation of pains and rewards in Williams' *Cases*—could be guides to personal identity. Whiting is mistaken to think that rejecting the independence thesis according to which personal identity obtains independently of subjects' concerns helps to address the charges developed in (4.3).

4.6 Debunking And Skepticism About Personal Identity

One worry about debunking arguments is that a successful local instance can explode into a global debunking argument that undermines way too many beliefs. For example, suppose that there is a successful evolutionary debunking argument against the belief that we are justified in extending altruistic attitudes only to persons in our vicinity.³⁵⁶ Suppose we grant the debunker that this belief has its causal origin in a non-truth-tracking process of evolutionary selection. The problem is that the very same worry seems to apply to any alternative belief, e.g., that self-interest or unrestricted altruism are justified. Presumably, these beliefs too would have the evolutionary history which the initial, local debunking argument took as off-track. In the worst case, we arrive at global evaluative skepti-

³⁵⁶Greene 2008, p. 76, as quoted by Kahane 2011, p. 113.

cism,³⁵⁷ or even skepticism beyond the evaluative domain.³⁵⁸ Assuming my local debunking argument against transplant intuitions succeeds, one might thus wonder whether it affects the justifiedness of all kinds of considerations about personal identity.

This worry can be resisted. Not all considerations about the topic result from transplant intuitions. Some of them are independent of the practical concerns that contaminate Locke and Williams' judgements. For example, one might prefer one criterion of personal identity over another on the basis of purely methodological considerations like parsimony,³⁵⁹ i.e. because it does not postulate the existence of more entities than we need for pursuing the explanatory aims of our metaphysics. Or we might take the idea that we trace people quite well in everyday life as a starting point, and maintain that only those accounts of personal identity are eligible that describe ourselves as entities whose persistence we are well-adapted to trace.³⁶⁰ The debunking argument is silent about the justifiedness of beliefs that result from methods other than transplant intuitions. This would include beliefs that might be *brought to our attention* through a transplant case, but subsequently receive *justifications* that are independent from practical concerns. The point of my discussion in (4.3) is that establishing such results requires much more argument by users of transplant intuitions.

³⁵⁷*Ibid.*, pp. 113-7 It is precisely the prospect of evaluative skepticism that motivates Street (2006, p. 141) to endorse anti-objectivism. Joyce (2000, p. 730) agrees with the dialectic, but opts for endorsing scepticism at least about moral beliefs, while also recommending that we retain moral norms as a useful fiction.

³⁵⁸For an overview, see Wielenberg 2016, section 6.

³⁵⁹For example Olson 2007, p. 36, 2015b, versus Parfit 2012.

³⁶⁰Johnston 1987, pp. 63, 77. My challenge is that we would need to ensure that it is indeed people, not practical concerns, that we are tracing well in everyday life. Moreover, I disagree with Johnston that his *human beings* fit his tracing desideratum.

My debunking argument actually helps to *respond* to a form of pessimism about criteria of personal identity. First, Sider argues that there is no fact of the matter whether the criterion is bodily or psychological.³⁶¹ He argues that talk about personal identity is indeterminate in meaning among the two views. There is no fact of the matter which one is correct. In a reasonably rich ontology, candidate referents for either criterion exist, but they are all equally eligible for being referents, e.g., by virtue of being equally *natural*. What Sider takes Williams to show is that our talk does not determinately pick out one candidate meaning. In *Case 1*, psychological persisters seem to be the referents, whereas it's a bodily persister in *Case 2*. The same inconstancy is frequent in ordinary talk about persons.

“A natural explanation is that ordinary thought contains two concepts of persisting persons, each responsible for a separate set of intuitions, neither of which is *our* canonical conception to the exclusion of the other.”³⁶²

Secondly, Eklund argues personal identity is an incoherent concept. He argues that conceptual competence involves a disposition to accept sentences that are *constitutive of meaning* of the expression or concept. In the case of personal identity, the sentences constitutive of meaning are jointly inconsistent. He motivates this picture by means of fission. Talk of personal identity has as a *meaning presupposition* that not more than one person at a time is continuous with someone at another time. If, as in fission, this presupposition is not satisfied, talk about personal identity *breaks down* by yielding inconsistent verdicts: continuity holds, so

³⁶¹Sider 2001a.

³⁶²*Ibid.*, p. 197.

there must be personal identity; yet asserting personal identity violates the transitivity of identity.³⁶³ Eklund suggests that Williams' cases reveal a further feature of our semantic competence: personal identity consists in *either* psychological *or* physical continuity.³⁶⁴ Eklund is a little unclear on what he thinks the relevant *meaning presupposition* is: that at most one person at a time bears either of the continuities to someone at another time, or that these continuities do not come apart, or both. Either way, Williams' *Cases* violate these presuppositions, and thus lead to perplexity.³⁶⁵

But according to both Sider and Eklund, nothing goes wrong when we employ transplant intuitions in Williams' cases. In fact, everything is going as it should: the intuitions reveal a feature of our talk or concept. The authors merely differ on the exact feature that is being revealed: Sider says that the meaning of talk about personal identity is to some extent *indeterminate*, whereas Eklund thinks the concept is *incoherent*.

It is worth highlighting that these claims concern the *meaning*, not the *truth values* of statements.³⁶⁶ It might be that Sider-Eklund-claims can exist *alongside* views where the facts to which terms and concepts refer, apply, or disapply determinately privilege on criterion of there other.

My debunking argument cautions against Sider-Eklund-style pictures. Transplant intuitions are bad evidence, and thus fail to justify personal identity judgements. The intuitions give no non-circular reasons to favour

³⁶³Eklund 2002, pp. 474-5.

³⁶⁴We might deny this with Sider: “[T]here are certain negative intuitions as well. [...] [M]y intuitions say not only that [in Case 2] *A* is the *A*-body-person, but also that *A* is *not* the *B*-body-person afterwards” (2001a, p. 200).

³⁶⁵Eklund 2002, pp. 480-1.

³⁶⁶*Ibid.*, p. 481.

one description over the other, and they are contaminated with practical concerns whose relevance to the topic is far from obvious or straightforward. If so, we cannot take features of this data as indicative of features of our talk or concept. The diagnoses of indeterminacy and inconsistency might be important, surprising, and correct. But we need more than transplant intuitions to support them.

4.7 Conclusion

Transplant intuitions motivate judgements on where one goes if one's psychology and body come apart. The best way of making sense of Williams-style diagnoses is that in such cases, the judgements take as evidence the structures of our person-related practices. I pointed at several problems with transplant intuitions. As is apparent in the *Prince and the Cobbler*, we can understand transplant intuitions either as offering illustrations of a theory or as attempts to convince us of a theory. In the former case, their dialectic aspirations are minimal. In the latter, they are unimpressive as they appear to provide only circular evidence, seem entirely uninformative if quasi-attitudes are possible, and are bad truth-trackers by virtue of diverging inexplicably across cases. Beliefs adopted on the basis of transplant intuitions thus can be debunked. This diagnosis is not limited to objectivism about personal identity. Neither does it motivate skepticism, just the demand to look for better evidence.

I close by highlighting three further implications. First, my findings deflate the relevance of empirical evidence on personal identity judgements when continuities come apart. For example, Nichols and Bruno

find that

“[i]ntuitions favoring a psychological continuity approach to personal identity are resilient across significant changes in the cases. [...] [I]ntuitions [at odds with psychological approaches] don’t seem to stand to scrutiny when examined in a broader and aggregative context.”³⁶⁷

In one sense, this fits Williams’ claims: presentational features do not seem causally responsible for our judgements. In another sense, there is a tension with Williams’ diagnosis of divergence: across the cases, intuitions favouring a psychological criterion are more resilient. Transplant intuitions thus seem less messy than Williams made it seem. However, if I am right, empirical evidence suggesting that transplant intuitions favour one theory over the other does not suffice to rescue their evidential value. Bad evidence does not get better by pointing resiliently in one direction.³⁶⁸

Second, Williams is the hallmark discussion for the worrisome, paradoxical content of our intuitions about personal identity. Our judgements shift frequently, leaning towards physical criteria in one case and psychological ones in the other—hence the Sider-Eklund diagnoses. Given the foregoing, we can indeed acknowledge that transplant intuitions as evidence about personal identity are in bad state, but not for the reasons standardly assumed. What makes them problematic is more than mere divergence and indecisiveness; it is their circularity and contamination with practical concerns pertaining to a different topic. Once proper evidence is taken into account, claims about indeterminacy or incoherence

³⁶⁷Nichols and Bruno 2010, p. 307.

³⁶⁸*Pace* Nichols and Bruno, I myself have found Williams’ divergent intuitions to be reproducible in seminar rooms with remarkable reliability. But as argued in the foregoing, this is besides the point.

might still be substantiated. For example, my points do not speak to Eklund's thesis that reflections on *fission* establish the incoherence of personal identity. But my discussion highlights the need to separate good from bad reasons for endorsing Sider-Eklund claims. Pointing at Williams' *Cases* is not enough.

Finally, animalists are right that transplant intuitions do not threaten their view. But I have provided new reasons for why this is so. It is not merely because animalism is consistent with transplant intuitions,³⁶⁹ or intuitions are unclear, and/or that it is not unreasonable to judge that we actually go with our bodies rather than our psychologies.³⁷⁰ Neither is there a need to argue that transplant scenarios are impossible.³⁷¹ Instead, transplant intuitions fail to support or justify hypotheses about personal identity, whether animalist or not. Nothing in my dismissal suggests that animalism is true, and does not face problems of its own. All I have argued is that transplant intuitions do not bring it or any other position under pressure. Scrutiny of the epistemic status of intuitions in favour of a theory can prevent us from endorsing the theory on the wrong grounds, and from rejecting others for no good reason.

³⁶⁹Olson 1997, ch. 3, Madden 2016.

³⁷⁰Snowdon 2014, pp. 209-10, 223, 236-7.

³⁷¹Williams 1956, pp. 242-4, Van Inwagen 1997.

Chapter 5

Against The Complex Versus Simple Distinction

As observers and disputants, we face choices about the theoretical tools by means of which we think about a given subject.³⁷² Sometimes, new concepts or distinctions advance discussion greatly. But there are other times when utilization of a concept or distinction blurs disputes or confuses us. I will now argue that the distinction between the complex and the simple view of personal identity is of this kind.

Debates on personal identity are often framed in terms of this distinction, and the literature acknowledges clear paradigm cases of either side.³⁷³ I examine three proposals on what the difference between the complex and the simple view is: Parfit's introduction of the terms (5.1), Gasser and Stefan's definition (5.2), and Noonan's recent proposal (5.3). I argue that given their shortcomings, the complex versus simple distinction fails to serve two important purposes of having a given distinction in our discourse: the facilitation of classification and endorsement of positions (5.4). We thus should not frame discourses in terms of the complex versus simple distinction.

The issue of delineation is pressing. One question is whether we should take the simple view seriously, especially because it is less frequently defended. Presumably, one minimum requirement is that a characterization of the view is available, ideally in terms of necessary and sufficient conditions³⁷⁴. Furthermore, it should be a characterization that classifies particular positions that are usually seen as paradigmatic

³⁷²The following material appears in Hummel 2017. Many thanks to two anonymous referees for helpful comments.

³⁷³Parfit 1982, p. 227. Garrett 2006, p. 122. Noonan 2011, p. 72.

³⁷⁴Compare Sider's criticism (2001b, pp. 63-8) that the credibility of Three-Dimensionalism is diminished because it can only be defined in negative terms, as *opposing* Four-Dimensionalism.

instances of the view *as* instances of the view. However, these are challenges not only for the simple view. After all, a further question is how informative the complex view is. If it opposes a view that cannot be clearly stated, the complex view would run the risk of being trivial or uninteresting rather than illuminating.

Throughout this text, I use the term 'view' as denoting what unites a family of positions, where a 'position' is a particular instance of a view. For example, S. Shoemaker's position is an instance of the complex view.

5.1 Parfit's Definitions

One straightforward way to investigate the difference between the complex and the simple view is to look in the place where the distinction is drawn for the first time.

5.1.1 Introduction of Complex/Simple And Reductionism/Non-Reductionism

Parfit defines:

“According to the Complex View, the fact of personal identity over time just consists in the holding of certain other facts. It consists in various kinds of psychological continuity, of memory, character, intention, and the like, which in turn rest upon bodily continuity. According to the Simple View, personal identity does not just consist in these continuities, but is a quite separate ‘further fact.’”³⁷⁵

In subsequent writings, Parfit drops the terms 'complex' and 'simple' without further announcement, and distinguishes reductionism from non-

³⁷⁵Parfit 1982, p. 227.

reductionism. Reductionists believe

“(1) that the fact of a person’s identity over time just consists in the holding of certain more particular facts, and
 (2) that these facts can be described without either presupposing the identity of this person, or explicitly claiming that the experiences in this person’s life are had by this person, or even explicitly claiming that this person exists. These facts can be described in an impersonal way.”³⁷⁶

In contrast,

“[o]ur view is Non-Reductionist if we reject both of the two Reductionist claims.”³⁷⁷

This rejection can happen in one of the following ways:³⁷⁸ by claiming that persons are

- separately existing entities, distinct from brains, bodies and experiences, for example Cartesian egos³⁷⁹;
- not separately existing entities, yet their existence is a further fact which does not just consist in physical and/or psychological continuity.

5.1.2 The Relation Between Complex/Simple And Reductionism/Non-reductionism

The relation between the complex view and reductionism on the one hand, and between the simple view and non-reductionism on the other, is not entirely clear. Parfit says that ‘complex’ and ‘simple’ are

³⁷⁶Parfit 1984, p. 210. He goes on to discuss several ways to enrich reductionism. See his clauses (3)-(9) in *ibid.*, pp. 211-2.

³⁷⁷*Ibid.*, p. 210.

³⁷⁸*Ibid.*

³⁷⁹One example is the *Featureless Cartesian View*. Here, the separate entity is unconnected to any observable facts; *ibid.*, p. 228.

“shorter labels for the two main views [reductionism and non-reductionism; P.H.],”³⁸⁰

but also explains:

“The *central claims* [my emphasis] of the reductionist tradition I shall call the *Complex View*, those of the non-reductionist tradition the *Simple View*.”

This leads some interpreters to regard the complex view and reductionism as equivalent.³⁸¹ Others think the complex view is a forerunner to reductionism.³⁸² Others think reductionism is a more specific doctrine than the complex view³⁸³ because of clause (2)³⁸⁴: the complex view is neutral on whether the more particular facts can be described without mentioning persons; Parfit's reductionism explicitly claims this.

So much seems clear: even though Parfit did not continue to talk in these terms, acceptance of the complex view is what unites the different reductionist positions, and the same goes for the simple view and non-reductionism. But he is not explicit on how much overlap there is between the complex view and reductionism and the simple view and non-reductionism respectively, for example whether they share their opinions solely on (1) or on (2) as well. For what follows, the answer does not really matter.

³⁸⁰Parfit 1982, p. 227.

³⁸¹Wasserman 2012, Fn. 1.

³⁸²Zimmerman 2012, p. 206.

³⁸³Noonan 1989, Ch. 5.6.

³⁸⁴Henceforth, I use italicized (1) and (2) to refer to Parfit's two conditions quoted above (1984, p. 210).

5.1.3 Parfit And Misclassifications

Parfit mentions paradigm cases of either side:³⁸⁵ Hume, Locke, Lewis and Parfit himself accept the complex view, Butler, Reid, Chisholm and Swinburne endorse the simple view.

This, I argue, should strike us as puzzling. The supposed paradigm cases of simplicity do not satisfy Parfit's definition. While Butler,³⁸⁶ Reid,³⁸⁷ and Swinburne³⁸⁸ are notoriously dismissive of positive stories about personal identity, they do operate against the backdrop of a Cartesian or Leibnizian metaphysics, and so I agree with Olson³⁸⁹ and disagree with Duncan³⁹⁰ that nothing in their works precludes acceptance of claims like the following:

- person x at t is the same person as y at t^* iff x and y have the same ego or soul;³⁹¹
- person x at t is the same person as y at t^* iff x is a monad, y is a monad, and x is the same monad as y .³⁹²

If so, they satisfy (1): a person's identity over time consists in more particular facts, namely that a certain soul or ego exists at the relevant times.

In Reid's case, there are two reasons why one might deny that (1) is satisfied. Firstly, Reid claims that persons are identical to monads,

³⁸⁵Parfit 1982, p. 227.

³⁸⁶Butler 1736, p. 100.

³⁸⁷Reid 1785b, pp. 108-9.

³⁸⁸Swinburne 1984, pp. 18-20.

³⁸⁹Olson 2012, pp. 49-50.

³⁹⁰Duncan 2014, pp. 284-5.

³⁹¹Given Swinburne 1984, p. 21.

³⁹²Given Reid 1785b, p. 21.

not just that personal identity consists in monad identity. Secondly, given that persons are identical to monads, facts about monads are not more particular than facts about persons. In response, I am prepared to accept that some simple positions do not satisfy (1). What matters for my purposes is that there are simple positions which definitely make consist-in claims and thus satisfy (1). Moreover, I deny that because of his identity claim, Reid should be seen as rejecting (1). Parfit explicitly includes positions that make identity claims between persons and other things under the umbrella of reductionism; he even calls them “hyper-Reductionist”³⁹³. In addition, positions according to which persons are identical to collections of matter or brains or biological organisms are certainly instances of the complex view.³⁹⁴ If these positions count as satisfying (1) by reducing facts about persons to more particular facts, then I see no reason why Reid could not do the same.

Moreover, authors like Swinburne and Reid can even satisfy (2): they can accept that the facts mentioned in (1) can be described in an impersonal way, without presupposing the existence or persistence of persons, for example by referring solely to egos, souls or monads. I claim that this holds even for Reid and others who think that persons are identical to these entities. This suspicion is motivated by the observation that no one takes a philosopher who believes that persons are identical to collections of material particles as being unable to provide an impersonal description of the world. In support, assume that persons are identical to collections of human-wise arranged, material particles. Suppose also

³⁹³Parfit 1995, p. 16.

³⁹⁴Gasser and Stefan 2012, pp. 4-7.

that I can describe the world exhaustively in terms of particle distribution. When describing the world, do I have to presuppose that persons exist? Of course, I can assume in advance that persons exist, and make my particle story reflect this. But I could also stay antecedently neutral on questions regarding persons, and proceed to tell you the exact distribution of each particle in a given universe. From this, it might or might not then follow that persons exist and persist. But if it does, it will not have something I have presupposed before making my description. And I suppose that exactly the same should apply to a philosopher who chooses to describe the world not in terms of particles, but in terms of monads, souls or egos. I am not claiming that there are passages in, e.g., Reid's writings which clearly suggest this reasoning. I also do not claim that he must follow this reasoning. But I do maintain that this reasoning is at least open to him and other simple positions.

Again, I do not claim that all simple positions accept (1) and (2). What matters for my purposes is that there are paradigm cases of simplicity which leave room for joint satisfaction of (1) and (2) and thereby satisfy the conditions for being complex. I conclude that Parfit has introduced a distinction, but not one that captures the paradigm cases.

5.2 Gasser And Stefan

One might hope that our grasp of the distinction has improved in the meantime. Gasser and Stefan in their 2012 volume on the complex versus simple distinction offer the following summarizing definitions:

“The complex view analyzes personal identity in terms of simpler

relations. The fact that a person persists over time is nothing more than some other facts which are generally spelled out in either biological or psychological terms, or both. That is, the complex view takes talk about what personal identity consists in literally. It aims to provide necessary and sufficient conditions for personal identity, thereby reducing it to the holding of basic biological or psychological relations. Whenever these relations obtain, personal identity obtains.

The simple view of personal identity, by contrast, denies that a person's identity through time consists in anything but itself. Biological and psychological continuity may be regarded as epistemic criteria for diachronic identity, but they are neither necessary nor sufficient conditions for personal identity. There are no non-circular, informative necessary and sufficient conditions for personal identity: personal identity consists in nothing other than itself."³⁹⁵

These definitions shift between different notions that are supposed to mark the difference: consisting in something other than itself, non-epistemic criteria for identity, definability, analysability, non-circularity, informativeness, necessary and sufficient conditions. These notions are certainly connected, but the connection is not straightforward enough to shift between them without further explanation. I will discuss them in turn.

Before I do this, let me add one clarification. By accepting the complex view as just defined, one need not accept that identity as such is analysable, or consists in other relations which are necessary and sufficient for its obtainment. One can still maintain that identity is a simple relation in the following sense:

“[i]dentity is utterly simple and unproblematic. Everything is identical to itself; nothing is ever identical to anything except itself. There is never any problem about what makes something identical to itself; nothing can ever fail to be. And there is never

³⁹⁵Gasser and Stefan 2012, p. 3.

any problem about what makes two things identical; two things never can be identical.”³⁹⁶

Here are two ways how: while identity relates one and the same thing to itself, features like those mentioned by Gasser and Stefan belong to a relation that connects distinct person stages³⁹⁷ or one person to more than one time.³⁹⁸ Whether or not one thinks that the latter relation is analysable in terms of other relations, or reduces to other relations, etc., the former can still be perfectly primitive, unanalysable, or ungrounded. The complex view can accept Lewis’ remarks on the simplicity of identity.

5.2.1 Analysability

As I understand analysability, it denotes a feature of either the meaning of a term or the content of a concept. In the first case, application of the term entails application of one or more further terms. In the second case, one or more constitutive principles of the concept are available. The meaning of ‘triangle’ entails, for example, ‘having three sides’. And the content of triangle is such that it has ‘applies to x if x has three sides’ as one of its constitutive principles. I see two reasons why rejection of an analysis of personal identity is not the mark of simplicity.

First, there is a sense in which paradigm cases of the simple view can be open to there being a term ‘personal identity’ or a concept personal identity that is analysable: they can accept an analysis of either in terms of, e.g., soul or monad identity.

³⁹⁶Lewis 1986, pp. 192-3.

³⁹⁷Lewis 1976, p. 20.

³⁹⁸Noonan 1989, p. 88.

Secondly, I do not have the impression that the disputants disagree on the status of a term or a concept. They are talking about what is the case at the object level, or the level of facts. The disagreement concerns the status of the things to which terms and concepts can refer, apply or disapply. If so, it is more promising to turn to metaphysical candidate claims.

5.2.2 Consist-In Claims

To understand the simple view as a denial of consist-in claims about personal identity leads to the problems described in (5.1.3): some paradigm cases of simplicity are not simple in this sense. If they accept (1), then they do defend consist-in claims about personal identity, although the more basic facts do not concern psychologies or bodies. I take it that analogous remarks apply to attempts to capture the difference in terms of idioms like ‘is nothing more’ or ‘reduces to’.

5.2.3 Necessary And Sufficient Conditions

Gasser and Stefan suggest that simple positions deny necessary and sufficient conditions for personal identity. But again, as described in (5.1.3), some simple positions do provide conditions for personal identity. And as I will describe in (5.3.2), I see no reason why simple positions should not be able to think of these conditions as necessary, sufficient or both. If so, they are not in disagreement with the complex view about the existence of such conditions.

5.2.4 Circularity/Informativeness

Are these conditions circular or informative? Of course paradigm cases of simplicity deem some paradigmatically complex positions circular. For example, Butler³⁹⁹ thinks this way about Locke's account. A more contemporary circularity charge is levelled by Lowe:⁴⁰⁰ he paraphrases Neo-Lockeanism as requiring that for person *p* to be identical to person *q*, any conscious experience had by *p* at any earlier time is quasi-remembered at least ancestrally by *q* at any later time, and vice versa. Lowe supposes it is important for Neo-Lockeanism that it is one and the same conscious experience which *p* and *q* share. He also supposes that any criterion for experiences will make reference to their subjects: *e* and *f* are the same experience if and only if *e* and *f* are qualitatively indistinguishable and had by the same person at the same time. If so, identity of experiences presupposes personal identity. Neo-Lockeanism is thus circular.

I am not concerned with whether Lowe's criticism is just. It is not obvious to me that Neo-Lockeans require that one and the same experience is shared by *p* and *q*. Two distinct mental states that stand in appropriate causal relations could be deemed sufficient, too. I would also like to remain non-committal on whether or not a given experience has its subject essentially.

What I would like to note is that although Butler and Lowe level circularity charges against Lockeanism, neither they nor simple positions in general need to deem any account of personal identity circular. For example, Lowe⁴⁰¹ explicitly denies that brain-based complex positions

³⁹⁹Butler 1736, p. 100.

⁴⁰⁰Lowe 2012, pp. 149-151.

⁴⁰¹*Ibid.*, pp. 151-2.

are circular (he rejects them for different reasons⁴⁰²): they take personal identity to consist in brain identity, and brain identity does not presuppose personal identity. Lowe seems to think that Swinburne's position is non-circular for analogous reasons.⁴⁰³ Thus, he deems only some but not all complex positions circular because their proposed criterion of personal identity in fact presupposes personal identity. And at least some simple positions are deemed non-circular because personal identity is defined in terms of entities of a different kind whose identity does not presuppose personal identity. If so, the existence of non-circular conditions of personal identity is not what the complex and the simple view disagree about.

Given (5.1.3), this is hardly surprising. We have seen that some simple positions can accept Parfit's condition (2), the possibility of a description of all relevant facts without presupposing the existence or persistence of persons. Such a description would contain non-circular conditions whose obtainment establishes personal identity. Moreover, I certainly would find it informative that personal identity consists in soul identity, and that I go where my soul goes. How I could have epistemic access to where my soul goes is a difficult question, but it concerns a different topic.

I conclude that Gasser and Stefan's account lacks precision and just like Parfit's account fails to capture the paradigm cases.

⁴⁰²*Ibid.*, pp. 142, 152.

⁴⁰³*Ibid.*, pp. 151-2.

5.3 Noonan

Noonan has recently provided a new proposal.

5.3.1 The Proposal

Noonan's proposal presupposes that problems of identity over time can be reduced to problems about kind-membership.⁴⁰⁴ For example,

“what makes a person P1 at a time t_1 identical with a person P2 at a time t_2 ?”⁴⁰⁵

can be rephrased as

“What conditions C satisfy the following schema: (P) If x is a person then if x exists at t and t^* , $Cxtt^*$?”⁴⁰⁶

The latter question does not mention identity. Instead, it concerns diachronic conditions on membership in the person-kind. Noonan prefers this reformulation since it avoids the impression that there are different identity relations for different kinds, or that identity can be reduced to other relations.

He suggests:

“the complex view is the view that *there are non-trivial, non-redundant, non-identity-involving diachronic constraints on personhood*. The simple view is that there are not, that the only

⁴⁰⁴Noonan 1989, section 5.3, crediting Quine 1976. For a denial that persistence conditions can be reduced to kind-membership, see Merricks 1998, section 4. If I understand correctly, Noonan and Merricks can agree that *some* candidate criteria for kind-membership do not contain diachronic constraints on the possible history of a member of the kind, and thus do not deliver persistence conditions. Merricks' discussion does not convince me that criteria for kind-membership *never* contain diachronic constraints; see also Noonan 2007, fn. 7.

⁴⁰⁵*Ibid.*, p. 123.

⁴⁰⁶*Ibid.*, p. 125.

non-trivial, non-redundant diachronic constraints on personhood are identity involving.”⁴⁰⁷

A diachronic constraint on personhood has the form:

“If x is a person, then if x exists at t and t^* , $Rxtt^*$.”⁴⁰⁸

Such a constraint is non-trivial if $Rxtt^*$ does not logically follow from the antecedent. And it is non-redundant if it is not entailed by the totality of synchronic constraints on personhood, each of which has the form:

“If x is a person, then if x exists at t , Fxt ”⁴⁰⁹

where F is a property of x that says nothing about any other times than t . Finally, a diachronic constraint on personhood is identity-involving if

“its satisfaction requires that something *other than* a person exists at times t and t^* ”⁴¹⁰.

Requiring the existence of one and the same soul at the relevant times would be an identity-involving constraint on personhood, provided persons are not souls. However, Noonan⁴¹¹ notes that Swinburne’s position fits this description and provides a criterion for soul identity over time:⁴¹² identity of form and continuity of immaterial stuff. Continuity of immaterial stuff does not require identity of immaterial stuff, and so Swinburne’s position includes a non-identity-involving constraint on personhood. Noonan’s proposal therefore makes Swinburne’s position count as complex. This classification is revisionary, but it is not what I will object to in the following.

⁴⁰⁷Noonan 2011, p. 74.

⁴⁰⁸*Ibid.*

⁴⁰⁹*Ibid.*, p. 73.

⁴¹⁰*Ibid.*, p. 74.

⁴¹¹*Ibid.*, p. 75.

⁴¹²Swinburne 1984, p. 27.

5.3.2 Refinement

Olson⁴¹³ argues that *pace* Noonan, simple positions accept non-trivial, non-redundant and non-identity-involving diachronic constraints on personhood. Assume with Noonan that any person is essentially a person, and consider the diachronic constraint

No Transformation (NT): if x is a person, then if x exists at t and t^* , it is not the case that x is wholly material at t and wholly immaterial at t^* .

This constraint is obviously not trivial or identity-involving. And in order to count as redundant, the diachronic constraint would have to follow from synchronic constraints on personhood. Olson thinks this need not be the case here. We can be neutral on whether

If x is a person, then if x exists at t , x is wholly material,

yet be convinced that *(NT)* is true. *(NT)* then would be a non-redundant constraint.

Now, since Noonan defines the simple view as the thesis that the only non-trivial, non-redundant diachronic constraints on personhood are identity-involving, a position that accepts *(NT)* is not simple. The trouble is, Olson says, that no philosopher denies *(NT)*. No one thinks that if you are now wholly material, you can become wholly immaterial.⁴¹⁴ Thus,

“Noonan’s proposal makes the simple view so strong that no one actually holds it”⁴¹⁵.

⁴¹³Olson 2012, pp. 56-7.

⁴¹⁴See also Duncan 2014, p. 286.

⁴¹⁵Olson 2012, p. 57.

This would be a bad result. The goal was to explain what unites proponents of the simple view, not to declare the simple view unendorsed. However, I see two reasons why Olson’s criticism can be resisted.

First reason: Some proponents of the simple view are not neutral on whether persons are material. Reid thinks that persons are monads, and monads are immaterial. He thus accepts (*NT*), but only as a redundant constraint, i.e., one that follows from synchronic constraints on personhood.⁴¹⁶ Depending on the particular position, similar remarks can apply if persons are taken to be egos or souls. Moreover, it is worth repeating that given (5.1.3) and (5.2.4), I deny that redundancy of this kind makes a position uninformative.

Second reason: Let us grant that (*NT*)-style constraints are not redundant. Olson is still exaggerating when he says that everyone accepts them. For example, Madell, a proponent of the simple view, writes:

“[W]hat unites the experiences in any of these biographies [of one and the same person across possible worlds] is that they belong, unanalysably, to the one mind, and that their doing so is compatible with the absence of objective connection between them.”⁴¹⁷

His suggestion seems to be that with respect to the possible history and life of a person, anything goes. The only restriction is that it needs to be her history and life, and this is a brute matter. Thus, Madell will reject (*NT*). And he is not alone: Chisholm⁴¹⁸ and Swinburne⁴¹⁹ offer similar considerations. These positions might seem bold, but it is not my intention to assess whether they are true. What matters for my purposes

⁴¹⁶A very similar point is made by Bottani 2013, p. 612.

⁴¹⁷Madell 1981, p. 107.

⁴¹⁸Chisholm 1969, p. 138.

⁴¹⁹Swinburne 1984, p. 25.

is that there are simple positions who reject, as Noonan⁴²⁰ recognizes, even the sparsest non-redundant constraints.

Olson can still complain that Noonan overgeneralizes from the fact that some or many proponents of the simple view reject (*NT*)-style constraints to the claim that all simple positions do so. If Olson shows that some accept non-redundant (*NT*)-style constraints without being subject to the First reason described above, Noonan's definition of the simple view would be too narrow.

But even then, I think we can improve Noonan's proposal accordingly. A natural thought is that if complex and simple views need not disagree on the features of available diachronic constraints on personhood, maybe they will differ on sufficient diachronic conditions

If x exists at t and x exists at t^* then $Rxtt^*$, then x is a person.⁴²¹

Presumably, we could define

*Simple**: If there are non-trivial, non-redundant, non-identity-involving diachronic constraints on personhood, $Rxtt^*$, then joint satisfaction of them does not guarantee that x is a person.

*Complex**: There are non-trivial, non-redundant, non-identity-involving diachronic constraints on personhood, joint satisfaction of which guarantees that x is a person.

*Simple** is compatible with there being

- (I) no constraints at all;
- (II) only constraints that are trivial, redundant and/or identity-involving
(Noonan's definition of simplicity)

⁴²⁰Noonan 1989, p. 94.

⁴²¹Cf. *ibid.*, p. 89.

as the conditional will be true if its antecedent is false. If (II), simple positions might even accept that these conditions are sufficient for personal identity. For example, if they think that only persons have souls, they could say:

if x exists at t and x exists at t^* , then if x at t and x at t^* have the same soul, then x is a person.

Of course, one difference in the sufficient conditions specified by the complex and simple view remains:

*Simple*_{continued}*: ...If there are sufficient diachronic conditions of personhood, these conditions are trivial, redundant, and/or identity-involving.

These modifications meet Olson's objection. I have granted Olson that simple positions can accept non-redundant constraints. But *Simple** accommodates Olson's supposition and constitutes only a slight tweak to Noonan's proposal.

5.3.3 Being Identity-Involving

The problem for Noonan's proposal is that being identity-involving is an unsuitable feature to focus on when defining complexity and simplicity. Consider a position according to which persons are not identical to biological lives, yet

Same Biological Life (SBL): if x is a person, then if x exists at t and t^* , x has at t the same biological life that x has at t^* .

Let us assume that the position accepts no other non-identity-involving constraints. Consider the judgements of Noonan's proposal if the position is extended to include one of the following further assumptions:

*There are no necessary and sufficient conditions for life identity over time.*⁴²² (SBL) requires the existence of something other than the person at both t and t^* : one and the same biological life. Given the assumption, nothing more can be said about when this condition is satisfied. Unlike in the Swinburne case, we cannot unpack a non-identity-involving constraint from (SBL). Noonan's proposal thus classifies the position as simple.

Mereological Essentialism. x at t can have the same biological life as x at t^* only if the life x has at t is composed of exactly the same parts as the life of x at t^* . Again, (SBL) requires the existence of something other than the person at both t and t^* : all of the parts that make up the biological life. No non-identity-involving constraint can be unpacked. Noonan's proposal classifies the position as simple.

However, without these further assumptions, the position presumably is complex: in order for x to exist at both t and t^* , likely some kind of continuity of parts rather than their identity is needed between the life x has at t and the life x has at t^* .

First of all, acceptance of (SBL) brings a position so close to paradigmatically complex positions⁴²³ that it is unclear why it should be labelled simple even in the wildest circumstances. It would be a stretch to accept this simply as a surprising feature of the proposal.

Secondly, whether or not a constraint is identity-involving turns on a relational property of the constraint: its being accompanied by certain other, external commitments. If extra theory tells an identity-involving story about the entities figuring in $Rxtt^*$, the position that accepts the

⁴²²Olson (2012, p. 57) also mentions this possibility. Why would anyone think this? One example: shrinkage cases like the brain transplants discussed in (chapter 4) are typically seen as causing problems for animalism. Supposedly, persons but not animals can be shrunk to brain or head size, so persons cannot be identical to animals. One amongst many potential responses considered by Snowdon is that the animal *can* survive being shrunk to the head. This, he claims, cannot be ruled out because "the unfortunate truth is that there is no agreed and established account of animal persistence" (2014, p. 237).

⁴²³For example Olson 1997, pp. 16-7.

constraint is simple (provided no other non-identity-involving constraints have been accepted); if not, it is complex (provided $Rxtt^*$ is neither trivial nor redundant). But no one has ever thought of the simplicity or complexity of a position as turning on a relational matter in this way. Neither does anyone think that a position turns from complex to simple or vice versa if it extends, contracts, or if its background theory changes. A nice anecdote from Olson⁴²⁴ is telling here: Swinburne nowadays thinks that contrary to what he used to believe, there is no criterion for soul identity over time. But no one, including Swinburne himself, thinks that his position turned from complex to simple. To be fair, no one except Noonan would have classified Swinburne complex in the first place. But suppose we accept (*SBL*), and later become convinced that, e.g., one of the two further assumptions above holds. According to Noonan, we first held a complex position and ended up with a simple one. Neither this diagnosis nor its grounds are anticipated in the literature.

5.4 Two Purposes Of A Distinction

We could stipulate to use terms in this way, but it is another question whether and why this is advisable. When we reflect on the problems of the examined proposals, it seems warranted to take a step back and ask: why are we interested in whether a view is complex or simple?

I see two minimal purposes of having a given distinction in our discourses: *Classification* and *Endorsement*. I will now explain them in turn.

⁴²⁴Olson 2012, Fn. 9.

Classification. A distinction can help us to classify positions with respect to shared features and to organize them into families of positions. This helps us to get an overview on their similarities and differences, and therewith facilitates navigation in the landscape of positions.

- (i) We can define a distinction and go through existing positions in order to determine whether or not they fall on one side of the distinction. Asserting that a position falls on one side of the distinction describes the position as satisfying certain conditions.
- (ii) Alternatively, we start out with a number of positions whom we suspect to share important features, group them provisionally, and then think further about what unites them. Given the clear picture of which positions are paradigm cases, this seems to be what we are doing in the case of the complex versus simple distinction.

Regarding *Classification*, Parfit's and Gasser and Stefan's definitions do not do a good job. Their conditions for complexity are jointly satisfied by paradigm cases of simplicity. There is a mismatch between the positions we antecedently group together and what follows from the definitions. Moreover, the definitions include imprecisions and shifts in key notions. Whether a given position lands on one side of the distinction depends on which notion the definition picks out. And we have seen that no specification gives us the right results (5.2).

Setting Swinburne aside, Noonan's account classifies extant paradigm cases reasonably well. However, its sensitivity to arguably irrelevant external commitments due to the relational nature of being identity-involving raises doubts about the grounds for these classifications. For the purpose of (ii)-style Classification, not just any feature shared by the positions will interest us. We are looking for features that are both shared by the positions and constitutive for each of them. Here is an analogy. Suppose we live in a world where all and only round things

happen to be red. It would be too quick to conclude that the difference between round and non-round things has something to do with redness. We should come to see that round things could be blue, transparent, etc., and so that facts about roundness can float free of facts about colours. The positions discussed in 5.3.3 are the analogues to these counterexamples. They show that even if most paradigm cases of complexity and simplicity respectively share the features Noonan's proposal mentions, serious worries remain whether possession of these features is distinctive of the respective view. Amongst others, we can thus remain sceptical whether the proposal can deal with the classification of new positions, and how it behaves once interpretation proceeds and gradually uncovers further commitments of a given author.

The difficulties encountered in the previous sections show that it is hard to see which features paradigmatic instances of complexity or simplicity share in the sense just described. The most charitable thing to say is that 'complex' and 'simple' try to stand in for more specific features. The problem is that it is not clear which ones these are. Given the failure of the examined proposals, there should be a significant amount of pessimism and no optimism about the availability of an account.

Classification of positions is not the only purpose of having a distinction:

Endorsement. A distinction can help us to express our thoughts through endorsement of one of the sides. Distinguishing one kind of view from another provides us with targets for affirmation and criticism. Again, there are at least two possible directions:

- (iii) We may find a number of positions independently plausible (or implausible). If we learn from Classification that they all fall on one side of a relevant distinction, and if we know of

no implausible (or plausible) position that falls on the same side, our affirmation (or criticism) can extend beyond the individual cases and to the side of the distinction as such.

- (iv) Alternatively, upon reaction on the distinction in abstraction from particular instances of either side, we become convinced that one side is correct, and therefore become sympathetic to its instances.

Suppose we endorse the complex view as in (iv). Because of the difficulties in *Classification*, this endorsement should strike us as odd. Complexity is due to the possession of features as specified in *Classification*. Given that this specification was unsuccessful, it is not clear what we have endorsed now. It is thus unclear on what grounds we made the endorsement and whether these grounds are reasonable. After all, we can hardly give reasons for the affirmation of a view if we cannot explain what the view is. Instead, (iv)-style Endorsement will only be plausible derivatively upon plausible affirmation of features we ascribe to the view.

Nothing seems wrong with (iii)-style *Endorsement* of the complex or the simple view. Indeed (iii)-style rejections of the simple view are common: it is often dismissed on the grounds that simple positions' stances on marks of complexity as defined by Parfit, Gasser and Stefan, and others are coherent but implausible.⁴²⁵ However, if what I argued is right, then these dismissals proceed on the wrong grounds. As shown in (5.1) and (5.2), simple positions can accept conditions which many take to be constitutive of the complex view. If so, further explanation and interpretation are needed to bring out what, if anything, is wrong with the simple view.

Further explanation is also needed on how supposed theoretical or

⁴²⁵S. Shoemaker 1985, p. 449. Duncan 2014, pp. 291, 293-4.

practical consequences follow from endorsement of either side of the complex versus simple distinction. For example, Parfit endorses the complex view and thinks it follows that identity is not what matters. Of course, he defends this latter claim not merely because he endorses the complex view (whatever that means), but because he thinks conditions like (1) and (2) obtain.⁴²⁶ He happens to think that acceptance of these conditions is what unites proponents of complexity. But this classification is not only unimportant, as the practical conclusions are thought to follow from conditions like (1) and (2) already; it is also mistaken. (1) and (2) can be accepted by simple positions. If there really is a difference in practical judgements between the complex and the simple view, more needs to be said on why this should be the case. (1) and (2) as well as the other conditions mentioned by the proposals cannot ground the asymmetry, because the complex and the simple view are not in disagreement about them.

Analogous remarks apply to the claim that the complex view cannot justify future-directed self-concern, or cannot justify it to the extent that the simple view can.⁴²⁷ Some take this as a *reductio* of the complex view and a reason to endorse the simple view,⁴²⁸ others as a reason for revision of ordinary patterns of concern that presuppose the simple view.⁴²⁹ But we have not yet seen a satisfying account of the difference between the complex and the simple view, let alone one that licenses these claims.

⁴²⁶Parfit 1984, pp. 260-3.

⁴²⁷*Ibid.*, pp. 307-12.

⁴²⁸Butler 1736, p. 102. Madell 1981, pp. 109-112. Swinburne 1974, p. 246.

⁴²⁹Parfit 1984, p. 280.

5.5 Conclusion

In spite of the elusiveness of what makes the difference, many philosophers are convinced that the distinction is an important one. Proponents of the complex view think that simple views are clearly wrong and that there are good reasons to endorse the complex view (1.1.2), while proponents of the simple view are convinced of the opposite (1.1.1). From the assumption that there is an important difference between complex and simple positions, authors have drawn conclusions about what matters, whether special concern for future selves can be justified, whether persons can undergo fission, and which theory of personal identity is true.⁴³⁰ But if I am right, taking sides in the complex versus simple distinction cannot play a role in these disagreements. This is because as shown, authors largely agree on the conditions discussed in the present chapter, yet disagree on the topics just mentioned.

Specifically, we get a new, straightforward reason for why we can discard claims according to which taking sides in the complex versus simple distinction should commit one to a certain stance about what matters: conditions like Parfit's (1) and (2) supposedly establish that contrary to what simple positions and common sense suggest, identity is not what matters. In response, not only can we mention authors who think that metaphysical insights, including insights on the nature of personal identity, concern the fundamental structure of reality and are thus neutral on ordinary facts of practical relevance⁴³¹ or that the relation between metaphysics and practical concerns is at least much less straightforward

⁴³⁰Parfit 1984. Swinburne 1984. S. Shoemaker 2012. Madell 1981; Madell 2015.

⁴³¹Sider 2013. Olson 2013.

than Parfit's revisionism suggests⁴³². And not only can we highlight familiar considerations of authors like Lewis⁴³³ who describe how a position can accept the conditions taken to be constitutive of the complex view and still be in perfect agreement with the proponents of the simple view regarding what matters and whether a person can divide. Given the present discussion, we can make a new point: conditions like (1) and (2) are satisfied by paradigmatically simple positions all along. Stances on these conditions alone thus cannot be what grounds differences in practical judgements.

In particular, we get another reason—besides (2.4)—for rejecting the idea that there is an asymmetry between the complex and the simple view in terms of how well they justify future-directed concern. As long as it is unclear what divides the complex and the simple view, and simple positions appear to be able to accept conditions thought to be constitutive of the complex view, we should be wary of the claim that complex but not simple positions can provide such justification. If the complex versus simple distinction is confused, we should also dismiss the idea that it marks a difference in reason-giving power.

Finally, the complex versus simple distinction turned out to be unhelpful with respect to *Classification*, and *Endorsement* should take place on the basis of more specific features anyways. If not in terms of complexity and simplicity, how should we frame debates about personal identity? I recommend speaking in terms of the specific conditions and related concepts for which the terms 'complex' and 'simple' try to stand in.

⁴³²D. Shoemaker 2007.

⁴³³Lewis 1976.

Promising candidate conditions presumably make positive or negative claims regarding the groundedness, criteriality, materiality, fundamentality, and/or the mereological status of persons and their persistence. It has been argued in this paper that the prospects for providing an account of the complex versus simple distinction are bad, and so we will likely arrive at a more fragmented landscape of views. But this fragmentation will be outweighed by benefits in clarity and precision. I will develop these suggestions further in the next chapter.

Chapter 6

Replacing The Complex

Versus Simple Distinction

Although many authors think that the complex versus simple distinction is important (chapter 1), I have argued that these labels are unhelpful (chapter 5). Most accounts of the distinction do not even get the most obvious paradigm cases right. Whatever unites the simple view is, as Olson puts it, “elusive”⁴³⁴. And I add that this also obscures what is distinctive of the complex view.

The question arises how else we should frame the debate. I will now sketch some positive, revisionary suggestions on this matter. My working hypothesis is that there is more than one issue at the heart of the complex versus simple distinction. I propose that if we want to compare, classify, and endorse theories of personal identity, we should employ notions that are more fine-grained and well-defined. Some fruitful questions include: does personal identity consist in something else? What does ‘consist in’ mean here? Are persons dependent on other things? If so, on what? Can we provide criteria for the persistence of persons? Are these criteria metaphysical or epistemic? Are persons fundamental constituents of the world? If personal identity consists in other facts, can we work out facts of personal identity from the fundamental facts?

My claim is not that stances on these questions exhaust the disagreement between complex and simple positions, and capture what *actually* is at the heart of the debate. The complex versus simple distinction turned out to be unclear. I thus demand that we should leave it behind. Instead, we should disambiguate and debate more refined questions and proposals. I will nevertheless continue to refer to the complex versus simple distinction in order to support my recommendations about what

⁴³⁴Olson 2012, p. 62.

should replace it. As I will make clear, I am not thereby offering analyses of, or proposals on, the distinction. Instead, I am merely illustrating that my suggestions for replacing the complex versus simple distinction fit *some* of the data from the paradigm cases. None of my suggestions for replacement capture the complex versus simple distinction neatly, which is exactly why I see them as *replacements* rather than analyses.

6.1 Ontological Dependence

It is often supposed that some things depend at all times of their existence on other things. This dependence is not merely causal, but ontological. It is not like the dependence of a football's flight trajectory on the particular kick it received. It is the kind of dependence that holds between the football and its atoms, the existence of a smile and the existence of a mouth that smiles, an event and its participants, non-empty sets and their members, tropes and their bearers, wholes and their parts, boundaries and extended objects, holes and their hosts,⁴³⁵ etc.

I will remain agnostic on the details of this relation of ontological dependence. Understood as *existential* dependence,⁴³⁶ x depends for its existence on the existence of a specific object y if (with ' E ' for existence, and \rightarrow for material implication)

Rigid existential dependence: $\Box(Ex \rightarrow Ey)$,

or on the existence F -type objects if

Generic existential dependence: $\Box(Ex \rightarrow \exists yFy)$.

⁴³⁵Correia 2008, p. 1015.

⁴³⁶*Ibid.* Tahko and Lowe 2016, section 2.

However, existential dependence can be seen as falling short of capturing ontological dependence. If x ontologically depends on y , it appears that x 's existence is somewhat derivative on, or less fundamental than, y 's existence. Here, existential dependence overgenerates derivativeness.⁴³⁷ For example, Socrates's existence necessitates the existence of the empty set: necessarily, if Socrates exists, the empty set does, too. But we hardly want to say that Socrates' existence is derivative upon the existence of the empty set. This is one motivation to consider further notions of dependence.

In some cases, we might speak of *essential* dependence:⁴³⁸ x would not be the object it is, had certain conditions not been met. For example, part of the essence of x (written as \Box_x) might be that a particular relation R obtains between x and y

$$\Box_x Rxy;$$

or x might be essentially such that it exists only if y does.

$$\Box_x (Ex \rightarrow Ey).$$

Moreover, following Fine⁴³⁹ we could deny that essential dependence can be fully captured in modal terms. Consider necessity and essentiality claims about Socrates and the singleton $\{\text{Socrates}\}$. Necessarily, if Socrates exists, then he is a member of the singleton $\{\text{Socrates}\}$. And necessarily, if $\{\text{Socrates}\}$ exists, then it has Socrates as its member. But there is an asymmetry in the corresponding essentiality claims. It seems plausible that $\{\text{Socrates}\}$ is essentially such that it has Socrates as a

⁴³⁷Correia 2008, p. 1023.

⁴³⁸*Ibid.*, pp. 1016-7. Tahko and Lowe 2016, section 4.

⁴³⁹Fine 1994.

member, whereas it does *not* seem part of Socrates' essence to belong to {Socrates}. This suggests that *essential* dependence does not reduce to modal dependence.

Besides *existential* and *essential* dependence, authors distinguish *explanatory* dependence. Correia formulates:⁴⁴⁰

$\Box(Ex \rightarrow Ex \text{ in virtue if the fact that/because/is explained by } Ey);$
 (Necessarily, if x exists, then this is in virtue of/because/is explained by the existence of y);

or for some feature G ,

$\Box(Ex \rightarrow \exists G(Ex \text{ in virtue of the fact that/because/is explained by } Ey))$
 (Necessarily, if x exists, then this is in virtue of/because/is explained by some feature of y);

Here, one might further distinguish different types of explanation by analogy with different types of modality, e.g., logical, conceptual, metaphysical, or nomological explanation, and consider some more salient than others for distinctively *ontological* dependence. For my purposes, the exact sense of ontological dependence will not matter.⁴⁴¹

Another issue on which I stay neutral is how ontological dependence relates to metaphysical grounding. They seem connected, but commentators point at differences.⁴⁴² In a sense, grounding is more demanding than ontological dependence. The former is typically seen as a *strict partial order*, and therewith as irreflexive, transitive, and asymmetric.⁴⁴³ In contrast, ontological dependence need not be asymmetric and can

⁴⁴⁰Correia 2008, pp. 1020-1.

⁴⁴¹For further variants of ontological dependence, see Koslicki 2012.

⁴⁴²Tahko and Lowe 2016, section 5.

⁴⁴³Raven 2013.

be reflexive, e.g., in its rigid existential variant.⁴⁴⁴ Moreover, ontological dependence need not be explanatory, whereas grounding typically is. And as just described, ontological dependence need not capture priority and fundamentality, while grounding typically does. At least for those varieties of ontological dependence that capture priority, Schnieder and Correia propose the following equivalences, depending on whether a predicational account of grounding (a relational predicate that connects facts) or an operational approach (a sentential connective that relates propositions) is preferred:⁴⁴⁵

Pred x ontologically depends on y iff some fact about y grounds the fact that x exists

Op x ontologically depends on y iff $\exists F(x$ exists because y is F).

If one accepts equivalences of this kind, then the following can easily be rephrased in terms of grounding.

In contrast to the examples above, some things do *not* depend for their existence on other things. Since Aristotle, a substance is often taken to be an entity that can exist on its own and depends on nothing else.⁴⁴⁶ For example, Descartes writes:

“[b]y substance we can understand nothing else than a thing which exists in such a way that it needs no other thing in order to exist”⁴⁴⁷.

The existence of mereological atoms can be seen as independent of other things. According to some notions of God, she would, if existing, not

⁴⁴⁴Correia 2008, p. 1023.

⁴⁴⁵Correia and Schnieder 2012, p. 25. For further discussion, see Schnieder [forthcoming](#).

⁴⁴⁶Correia 2008, pp. 1025-6 and Tahko and Lowe 2016, section 6.3 argue that mere *existential* dependence is unsuitable for characterizing substances, and thus recommend to focus on *essential* dependence.

⁴⁴⁷Descartes 1647, p. 210.

depend on any other things. According to Schaffer, the world depends on nothing, not even on its parts, for they depend on the world.⁴⁴⁸ This is an instance of *holism* understood as “the doctrine that the parts of a whole can depend upon the whole itself.”⁴⁴⁹

Are persons ontologically dependent? Such claims do seem to play a role in the simple view. Here are three examples. First, in response to Locke’s suggestion that personal persistence consists in sameness of consciousness, Butler objects that Locke confuses the phenomenon with its evidence. The fact that a series of conscious moments belongs to the same person cannot constitute a person’s persistence; it presupposes it. As quoted earlier (1.2.2):

“one should really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in any other case, can constitute truth, which it presupposes.”⁴⁵⁰

We can interpret Butler as accusing Locke of a mistake about *explanatory* ontological dependence. There is one and the same consciousness over time *because* the person persists. Moreover, he comes close to claiming the ontological independence of persons: a person is either a substance or the property of a substance.⁴⁵¹ In the latter case, the person might not be ontologically independent, yet she would be the property of something else that is.

Second, Reid claims that “a person is something indivisible, and is what Leibniz calls a *monad*”⁴⁵², i.e. an indivisible, ontologically inde-

⁴⁴⁸Schaffer 2010.

⁴⁴⁹Fine 1995, p. 270.

⁴⁵⁰Butler 1736, p. 100.

⁴⁵¹*Ibid.*, p. 104.

⁴⁵²Reid 1785b, p. 109.

pendent, fundamental constituent of the world. In particular, attempts to characterize persons in terms of other notions are mistaken:

“I am not thought, I am not action, I am not feeling; I am something that thinks, and acts, and suffers.”⁴⁵³

Again, part of the complaint is that others have reversed the relation between personal identity and what it supposedly depends on.

Third, Lowe claims that persons, or *selves* as he calls them, are *simple* substances.⁴⁵⁴ He has in mind a distinctive kind of simplicity: *selves* might possess *spatial* parts, but possess no *substantial* parts,⁴⁵⁵ i.e. parts that are themselves substances. Since Lowe is convinced that reference to substantial parts is necessary for the formulation of persistence criteria, and denies that selves have substantial parts, he concludes that there are no persistence criteria for selves. He goes on to deny that this makes *selves* mysterious or uncaused. Continuous brain functioning can be accepted as a necessary condition for the existence of embodied *selves*, natural laws can set further constraints on possible histories, and inner lives of *selves* need to be at least minimally coherent.

“But it would not follow from this that the identity of the self over time is grounded in continuity of brain-function, or indeed anything else.”⁴⁵⁶

No definite conditions for the coming- or ceasing-to-be of *selves* can be given. Evidence about brain functioning or lack thereof is inconclusive about whether a *self* exists. One way to understand Lowe’s view is that

⁴⁵³Reid 1785b, p. 109.

⁴⁵⁴Lowe 1996, pp. 39-44. Like the authors above, he roughly means an Aristotelian *primary* substance, i.e. a concrete individual thing, or *continuant*; cf. *ibid.*, pp. 2-3.

⁴⁵⁵*Ibid.*, p. 36.

⁴⁵⁶*Ibid.*, p. 41.

as a matter of contingent fact, selves *causally* depend on things like brain functioning, yet remain *ontologically* independent and ungrounded.

I suspect that in the end, we will *not* find that all complex positions are committed to the ontological dependence of persons, while their opponents claim the contrary. Many complex positions are not sufficiently characterized by claims of ontological dependence. They do not only think that persons are dependent, but also that metaphysical criterialism is true, i.e. that there are conditions individually necessary and jointly sufficient for a person to persist. It is not obvious how this additional claim can be captured in terms of dependence.⁴⁵⁷

Ontological dependence is also broader than Parfit's condition (1) and (2)—the availability of consist-in claims and complete impersonal descriptions—which he thinks are constitutive of the complex view (5.1.2). If x -entities are reducible to y -entities, then x -entities depend on y -entities. I cannot think of a case of reduction where the reduced entity is independent of the reducing entity/entities. The converse does not hold; if x -entities depend on y -entities, x -entities need not be reducible to y -entities. Moreover, if x -entities are not reducible to y -entities, this does *not* mean that x -entities do not depend on y -entities. Because of these differences, we should avoid distorting positions and be clear about whether they speak of reduction, dependence, or both.

While the complex view is not fully captured by means of dependence claims, I argued earlier that the simple view need not deny the consist-in claims about persons and their persistence, not even the availability of informative criteria (5.1.3, 5.2.2, 5.2.3). Amongst others, I noted that

⁴⁵⁷See also Olson 2012, p. 49.

some simple theorists take persons to depend on souls or egos, or even *identify* persons with monads. I argued that this need not amount to a denial of consist-in claims. We can now point at another possibility for how consist-in claims could be made alongside identity claims about persons and, e.g., monads. If the consist-in claim takes the form of a claim of ontological dependence, it can be reflexive, e.g., when it comes in the *existential* variant. Here, we would have the expressive resources to say *both* that persons are *identified* with monads *and* that persons *depend* for their existence and/or identity on the monads they are identified with. This gives another way of articulating how the simple view can accept conditions normally thought to be constitutive of the complex view. Simple positions can accept the ontological dependence of persons.⁴⁵⁸

I thus conclude that ontological dependence is useful towards understanding the commitments of complex and simple positions. This is true even though it is not the case that the simple view denies the ontological dependence of persons, and the complex view accepts it.

One challenge to phrasing complex and simple positions in terms of their stance on ontological dependence is sometimes raised against metaphysical grounding.⁴⁵⁹ Apart from controversies about, e.g., the logic of ground or particular instances of grounding, more general worries concern the intelligibility, unity, instantiation, and/or epistemic status of grounding claims. Some authors report that they do not understand what grounding means.⁴⁶⁰ Others think that the phenomena subsumed

⁴⁵⁸Earlier, I also noted that for more extreme variants of the simple view (such as Madell 1981), anything goes. In the following (6.3), I provide further suggestions on how to capture these views.

⁴⁵⁹Correia and Schnieder 2012, pp. 29-30.

⁴⁶⁰Daly 2012.

under the notion of grounding are way too diverse to be seen and systematically explored as instances of one and the same relation.⁴⁶¹ Moreover, we already have a battery of metaphysical and logical notions, e.g., supervenience, possible worlds, entailment, and it is not obvious what grounding adds.⁴⁶² Similar worries could be formulated about ontological dependence.

If grounding and dependence are confused notions, then it might be misleading or uncharitable to express complex and simple positions in terms of it. But my suggestion also offers an opportunity. Personal identity could turn out to be a discourse in which these notions achieve what the complex versus simple distinction could not: clarifying their commitments and thereby facilitating comparisons, *classifications* and *endorsements* of positions (5.4). In this case, it would not be necessary that meta-metaphysics approves *ex ante* of ontological dependence or grounding in order for them to become eligible to figure in debates. Usefulness in the first-order discourse could motivate a presumption of legitimacy for these notions.

We need not endorse full-blown skepticism about dependence and grounding to note that if they play a role in the complex versus simple distinction, the debate could turn out to be more intricate and puzzling than participants assume. There is more than one sense of ontological dependence, and it is difficult to pin down the most salient and straightforward meaning, let alone to provide an analysis or reductive definition of the notion. In tension with the emphatic endorsements and rejections

⁴⁶¹Hofweber 2009.

⁴⁶²Rosen demands “ideological toleration” (2010, p. 109), whereas Wilson (2014) argues that there is no need for grounding.

of the complex and the simple view (chapter 1), at least one central metaphysical notion at issue is potentially primitive and less straightforward than authors might have expected. For example, part of the appeal of the complex view is that it offers an analysis of personal identity, improves our understanding of persons, and refutes “Identity Mystics”⁴⁶³. But if ontological dependence plays a role in these projects, then they build on a notion that could be less transparent than proponents of the complex view take their own positions to be. Further questions about the exact sense and nature of the relevant dependence relation arise. Such complications can be seen as a reason to refrain from introducing ontological dependence in the complex versus simple debate. But alternatively, they might make us realize that certain aspects of the supposed opposition and the disputants’ self-understanding need to be revisited.

6.2 Fundamentalism About Persons

Some metaphysicians distinguish fundamental from derivative entities. Fundamental are those things from which the world is made up, i.e. whose creation would suffice to make a world perfectly like ours. Derivative things are those which are made up from the fundamental ones. Although derivative things exist, fundamental things exist *in reality*.⁴⁶⁴ Barnes provides the following glosses:⁴⁶⁵

God’s creation metaphor: Fundamental are those things that God needs to create in order to make the world how it is.⁴⁶⁶

⁴⁶³Zimmerman 1998, p. 281.

⁴⁶⁴Fine 2001.

⁴⁶⁵Barnes 2012, pp. 875-8.

⁴⁶⁶This picture comes from Kripke 1972, pp. 153-4.

Non-fundamental (i.e. derivative) things are somewhat superfluous in a description of a world.

Truthmaking: Fundamental are those things which truthmake their own existence, and which are capable of serving as truthmakers for the existence of other, derivative entities.

Non-redundant causal roles: Fundamental are those things that cause events and sustain other things, and do so non-redundantly.

Some authors accept only two levels, the fundamental and the derivative, whereas others are convinced that there needs to be more: a lowest level of facts and a plurality of further, hierarchically ordered levels, with the fundamental at one end.

According to one understanding of fundamentality, it relates to ontological independence and grounding.⁴⁶⁷ The most fundamental level is comprised of ontologically independent or ungrounded entities. Derivative entities depend on, or are grounded in, other entities that are either fundamental *simpliciter*, or *more* fundamental and/or *less* derivative. However, in the preceding section, I mentioned some differences in the formal properties of ontological dependence and grounding: the former can be reflexive and need not be asymmetric, the latter is typically seen as a strict partial order. For similar reasons, it can be disputed that ontological dependence and the less-fundamental relation go together, as the latter is irreflexive and asymmetric.⁴⁶⁸

In fact, Barnes suggests that the fundamental versus derivative distinction and the independent versus dependent distinction cut across each other. For ontological dependence, she suggests the following litmus test:⁴⁶⁹

⁴⁶⁷Tahko and Lowe 2016, section 6.4.

⁴⁶⁸Correia 2008, p. 1023.

⁴⁶⁹Barnes 2012, p. 880.

(OD) An entity x is dependent iff for all possible world w and times t at which a duplicate of x exists, that duplicate is accompanied by other concrete, contingent objects in w at t .

Barnes also assumes that anything which does not satisfy (OD) counts as ontologically independent. For example, mereological atoms can exist without being accompanied by other things, and are thus independent.⁴⁷⁰ Most independent things are fundamental. For example, mereological atoms do need to be set in place when creating the world, and the same goes for other examples for ontological independence from the previous section. Moreover, most things that satisfy (OD), such as mereologically complex entities,⁴⁷¹ are derivative—they do not require extra steps in the making of the world; their existence is secured once the fundamental things are in place.

But Barnes argues that there is conceptual room for more: first, some entities are independent but non-fundamental. On some accounts of mathematical truths,⁴⁷² they do not impose constraints on what the world is like, and thus do not require additional truthmakers or steps in the creation of the world. Thus, they could count as non-fundamental. At the same time, they satisfy (OD), and thus combine derivativeness with ontological independence. Second, some entities are fundamental but dependent; Barnes calls them *emergent*.⁴⁷³ For example, some think that minds and mental properties depend on physical properties, but

⁴⁷⁰For the sake of presentation, I will presuppose that there are mereological atoms, i.e. that the world is not gunky. If the world was gunky, we would have *even more* reason to accept Barnes' claims about *emergence*; see Barnes 2012, pp. 888-9.

⁴⁷¹Unless one endorses the Schaffer picture.

⁴⁷²Barnes is thinking of Rayo 2009.

⁴⁷³Barnes 2012, p. 884.

have distinctive causal powers. Something needs to be added to the microphysical arrangement of the world in order for there to be minds, even though minds satisfy (OD) as they are caused and sustained by the physical. Barnes is not defending these pictures, but argues that *if* they are accepted, they would provide instances where dependence and derivativeness on the one hand, and independence and fundamentality on the other hand come apart.

One example for emergent entities in Barnes' sense are persons in the ontologies of Van Inwagen and Merricks.⁴⁷⁴ They claim that mereological composition occurs only for living organisms and persons. Barnes paraphrases:

“Simples (arranged object-wise) can do all the work we need for ordinary macroscopic objects. But they cannot do all the work, the thought goes, for living beings (van Inwagen) or more specifically for persons (Merricks). These entities have special properties, causal powers, etc., that simply cannot be accounted for solely with reference to simples.”⁴⁷⁵

Merricks' motivations are as follows:⁴⁷⁶ contrary to folk ontology, events in the macrophysical world are caused by mereological atoms acting in concert. Moreover, events are not overdetermined; they do not have more than one independent cause. Thus, if there were macrophysical objects, they would be causally irrelevant or *epiphenomenal* to what their constituent atoms are causing. Merricks goes on to argue that causally inert, epiphenomenal material objects should be *eliminated* from one's ontology, given that there remains no good reason for believing in their existence. However, human persons *are* capable of causing more

⁴⁷⁴[Ibid.](#), pp. 887-8.

⁴⁷⁵[Ibid.](#)

⁴⁷⁶Merricks [2001a](#), Ch. 3.

than what their constituent parts are causing. They possess conscious mental properties that do not supervene on what their parts are like.⁴⁷⁷ And human persons cause events partly by having such non-supervenient mental properties. There is thus no redundancy in the composite person's causing events in the macrophysical world. Eliminativism about persons is unmotivated.⁴⁷⁸ Persons are full-fledged ontological constituents of the world, possess special causal powers and have properties that cannot be accounted for in terms of the mereological atoms that compose them.

It is tempting to read participants in the complex versus simple debate as making claims about fundamentality. For example, Chisholm argues against the suggestion that persons are *entia per alio*, or “ontological parasites”⁴⁷⁹, which possess properties only derivatively, in virtue of being related to other, fundamental entities that possess properties non-derivatively. Reid claims that persons are identical to monads and therewith fundamental constituents of the world. In Barnes' terms, they must be put in place to make the world how it is. Moreover, remember Madell's extreme, simple position according to which a person's survival is compatible with *any* kind of discontinuity. This is plausibly understood as a fundamentality claim: once all the physical, psychological, and other continuities are in place, it is not settled yet whether a person survives. An extra truthmaker or step in the creation of the world is needed.

In contrast, Parfit's clauses (1) and (2)—the availability of consist-in claims and complete impersonal descriptions—suggest that for him, only

⁴⁷⁷Merricks 2001a, pp. 88-9.

⁴⁷⁸*Ibid.*, p. 113.

⁴⁷⁹Chisholm 1976, p. 104.

the *more particular* facts are fundamental. Once facts about brains, bodies, and psychologies are settled, we know all there is to personal identity.⁴⁸⁰ They are the truthmakers for personal identity. No extra steps are needed. Person facts are merely derivative. This also fits Parfit's comments that the simple view believes in *deep* further facts of personal identity, whereas the complex view denies their special metaphysical depth.⁴⁸¹

Fundamentality claims do seem to play a role in the disagreements. It is thus useful to rephrase their positions in these terms. At the same time, I deny that the fundamentality of personal identity marks the divide between the complex and the simple view. The latter can deny that personal identity is fundamental—for analogous reasons why they can accept consist-in claims like Parfit's (1). For example, consider Swinburne's position that persons have a soul and a body, but go with their souls if body and soul come apart. Assume further that there is no criterion for soul identity over time, just in case Noonan (5.3) has convinced you that this matters. Swinburne can agree that body and soul facts are fundamental; they are needed to make the world the way it is. But note that once these fundamental facts are given, person facts *derive*. Nothing more is needed for them to obtain.

As I argued earlier, even in Reid's case there is a question about how to understand the identity between persons and monads. I suggested that Reid could give a full description of the world in terms of monads, and as

⁴⁸⁰There are conceivable cases in which it is indeterminate whether our person concept still applies, but “[s]uch questions are, in the belittling sense, merely verbal” (Parfit 1995, p. 25).

⁴⁸¹Parfit 1984, pp. 262, 279-82, 309.

a consequence settle the person facts without having presupposed them—just like a materialist can give a material description of the world without presupposing person facts *and at the same time* accept that person facts *just are* facts about material wholes (or whole-wise arranged atoms). Within Reid's picture, we can say that once we have put the monads in place, the person facts are settled. But note that this gives a sense in which person facts are not fundamental. No extra truthmakers beyond monad facts are needed.

Not only can the simple view accept that personal identity is derivative; the complex view can see personal identity as fundamental. Suppose Reid clarifies that my suggested reading is *not* what he has in mind. Persons are monads, and monads are fundamental building blocks of the world. A complete description of the world in terms of its fundamental facts must mention persons. Now, compare Reid's position with animalism according to which persons are identical to human animals. Suppose we take human animals, if not themselves fundamental constituents of the world, as identical to collections and arrangements of mereological atoms. For the same reason that Reid makes a fundamentality claim about persons by identifying them with some of the world's fundamental building blocks, this version of animalism would, *pace* Parfit, make a fundamentality claim about persons by virtue of identifying them with collections and arrangements of fundamental constituents of the world. The fact that paradigmatically complex positions can accept fundamentality claims gives another reason for doubt that they are the mark of simple view.

Nevertheless, phrasing positions in terms of fundamentality can be illuminating. First, just like ontological dependence, the notion is more informative than the ambiguous and unclear labels ‘complex’ and ‘simple’. It would improve our understanding of a position’s commitments. Second, it cautions against erroneous *endorsements* of either the complex or the simple view. A prior commitment to the fundamentality of personal identity might appear to be a good reason to endorse the simple view. But we have just seen that fundamentality can be had within the complex view, too. The commitment thus should not be a knock-down argument against the complex view. At the same time, a prior commitment to the idea that personal identity depends on other relations, and maybe can be fully explained in terms of them, suggests it is derivative rather than fundamental, and this might appear to be a good reason to endorse the complex view. But we saw that fundamental entities too can depend on other entities, namely if they are *emergent* in Barnes’ sense. The prior commitment to the idea that personal identity is derivative upon other relations thus does not decisively favour the complex view. It can also be captured by the simple view.

6.3 Essentialism And Haecceitism

According to *essentialism*, at least some of an object’s properties are essential. Roughly, a property is essential if the object could not exist without it, and/or if the property contributes to making the object what it is.⁴⁸² If not, the property is accidental. The object could have failed

⁴⁸²Roca-Royes 2011, pp. 65-7.

to have it and yet exist and/or be the object it is. Essentialism puts constraints on an object's *de re* possibilities, i.e. the properties it could possibly have.

Some alleged essential properties are *trivial*⁴⁸³ in that every object has them. Examples include *being self-identical* or *being F-or-not-F*. Others are substantive, or *non-trivial*. For example, Kripke defends the idea that a person's biological origins are essential to her, i.e. she could not exist without stemming from the particular sperm-egg pair from which she results in the actual world.

Haecceitism is, roughly,⁴⁸⁴ the view that two worlds can be qualitatively indiscernible but differ non-qualitatively. The qualitative versus non-qualitative distinction has its own complications.⁴⁸⁵ Roughly, non-qualitative properties make reference to particular individuals (such as *being Napoleon*), whereas qualitative properties (such as *mass* or *charge*) do not. While essentialism constrains *de re* possibility, haecceitism extends it⁴⁸⁶ by allowing that non-qualitative matters can float free from qualitative matters.

Haecceities, or individual essences, are necessary and sufficient conditions for being a certain individual. If Kripke's essentiality of origins is accepted not only as a necessary but also as a sufficient condition, it would offer one account of persons' haecceities. *Extreme haecceitism* holds that the only essential properties are the trivial ones, such as *being*

⁴⁸³Roca-Royes 2011, p. 66.

⁴⁸⁴As Cowling (2016, section 1) explains, it is challenging to state haecceitism in a theory-neutral way.

⁴⁸⁵Diekemper 2009 and Cowling 2015 argue that the distinction must remain primitive.

⁴⁸⁶Cowling 2016, section 3.

self-identical. In other words, this view is extremely permissive on the ways an object could be.

Note that acceptance of *haecceities* does not entail acceptance of haecceitism. This is because the notion of a haecceity is neutral on the extent to which *qualitative* properties figure in it. If one takes *haecceities* to supervene on qualitative properties, one *denies* haecceitism; worlds could not differ non-qualitatively without differing qualitatively. However, many proponents of individual essences characterize them in impure or hybrid terms, i.e. in terms of both qualitative and non-qualitative aspects. If an individual figures irreducibly in another individual's *haecceity* or individual essence, a relation of *essential* ontological dependence as described earlier (6.1) results.⁴⁸⁷

Cowling suggests that haecceitism is useful towards characterizing the simple view.⁴⁸⁸ He understands the simple view as suggesting that personal identity is not reducible to non-qualitative properties. This amounts to a form of haecceitism.⁴⁸⁹ One observation in support of Cowling's reading is the simple view's stance towards fission. Chisholm⁴⁹⁰ and Swinburne⁴⁹¹ believe that the physical and psychological continuity relations in fission are not more than fallible evidence for personal identity. Strictly speaking, the authors deny that there are genuine fission cases. Persons never divide. In what appears to be fission, they go with only

⁴⁸⁷Roca-Royes 2011, p. 72.

⁴⁸⁸Cowling 2016, section 7.2.

⁴⁸⁹In line with the suggestion discussed in the preceding section, Cowling (*Ibid.*, section 2) clarifies that this haecceitist presupposition connects with a fundamentality claim: qualitative properties do not suffice to fix all the facts; some non-qualitative properties must be fundamental.

⁴⁹⁰Chisholm 1976, pp. 111-2.

⁴⁹¹Swinburne 1984, pp. 18-21.

one of the fission products. Facts about continuities fall short of determining which one it is. Two fission cases that are exactly alike in terms of continuity relations can differ in personal identity facts, in particular which fission product is identical with the fissioning individual. In other words, there can be qualitatively indiscernible fission scenarios that differ in personal identity facts. One explanation Swinburne gives is that a simple theorist can regard persons as mereologically simple, and thus as incapable of dividing.⁴⁹²

Similarly, Parfit's Featureless Cartesian View and Madell's position are easily read as endorsements of haecceitism. Parfit writes:

“[o]n [the Featureless Cartesian] view, history might have gone just as it did, except that I was Napoleon and he was me. This is not the claim that Derek Parfit might have been Napoleon. The claim is rather that I am one Cartesian Ego, and that Napoleon was another, and that these two Egos might have occupied each other's place.”⁴⁹³

Another way to put these positions is that they accept only a minimal conception of which properties are essential to an individual: only non-qualitative ones that make specific reference to the identity of their possessor, such as *being Parfit* or *being Napoleon*. And instantiation of these properties is a brute matter, or as Madell puts it, “compatible with the

⁴⁹²Swinburne 1984, p. 21 However, it is not obvious that mereological simplicity precludes division. For example, Hawley (2004, p. 393) suggests that a mereologically simple entity could gain parts by developing sufficient internal structure. Consider also J.R.G. William's proposal of higher-level universals, e.g., *being a proton**, which can hold of a particular without necessarily instantiating the particular's actual parts, e.g., quarks in the actual world. This allows simple instantiators in some worlds, e.g., in “a world qualitatively like our own ‘from the protons up’ but where there are no correlates of our quarks, but where the things which play the proton-role are mereologically simple” (2007, p. 199). These examples show that the move from mereological simplicity to necessary indivisibility is less straightforward than Swinburne and others realize.

⁴⁹³Parfit 1984, p. 228.

absence of objective connection.”⁴⁹⁴ In one sense, these positions are extremely permissive: a person can be *anything*. In another sense, they are extremely demanding: for any particular qualitative—e.g., psychological or bodily—makeup, it is an open question who this person is.

Parfit deems the Featureless Cartesian View “unintelligible”⁴⁹⁵. Now, these positions certainly provide only a very limited basis to ascertain and settle disputes about personal identity, given that observable or qualitative facts are deemed irrelevant. And they provide a curious combination of permissiveness and demandingness. But Parfit’s charge of unintelligibility is nevertheless misguided. These positions are perfectly intelligible. As with ontological dependence and fundamentalism about persons, phrasing them in terms of essentialism and haecceitism is illuminating and makes their commitments precise. It is just that the positions’ contents might strike some as surprising and worth rejecting.

It is also tempting to read Butler’s circularity objection as making a point about non-qualitativeness. Butler argues that sameness of consciousness cannot be what personal identity consists in. Instead, sameness of consciousness presupposes personal identity. We might rephrase: the sameness of consciousness requires something more than qualitative sameness. In order to be sure that a consciousness at one time is the same as a consciousness at another time, something non-qualitative needs to be known, such as *whose* consciousness it is. And this knowledge or entitlement is unavailable while trying to give an account of personal identity. Otherwise, the account would lapse into circularity. And in response, S.

⁴⁹⁴Madell 1981, p. 107.

⁴⁹⁵Parfit 1984, p. 228.

Shoemaker's quasi-memory can be understood as a kind of psychological sameness that eschews these non-qualitative features.

However, it is not the case that the simple view *per se* endorses haecceitism, and the complex view denies it. Things are more complicated. Consider once again Swinburne's position. Cowling phrased its stance towards fission as allowing that two scenarios can differ in terms of who is who while being qualitative exactly alike. It all depends on where the soul goes. But note that Swinburne need not believe that where the soul goes is a non-qualitative matter. It is at least open to him and other simple theorist to believe that soul facts or properties can be characterized without referring to individuals, e.g., without making claims about *whose* soul goes where. They could treat soul properties on par with other paradigmatically qualitative properties like mass and charge. If so, the theory's stance on the two fission scenarios would not be an instance of haecceitism; the scenarios would be qualitatively discernible. While Featureless Cartesianism and Madell are properly described as haecceitist, this version of Swinburne's position brings out that the simple view need not commit to haecceitism.

A related suggestion is that through clauses like Parfit's (1) and (2), the complex view is committed to the availability of qualitative analyses of non-qualitative properties like *being identical to Napoleon*. This property is paradigmatically non-qualitative by making reference to a particular individual. But Parfit's (1) and (2) state, and Featureless Cartesianism and Madell deny, that qualitative conditions can be specified which guarantee that the non-qualitative property is instantiated. Given the foregoing, simple positions that are less extreme and instead compat-

ible with the Swinburne-style picture could agree that non-qualitative properties can be described in this way.

I conclude that just as with ontological dependence and fundamentalism, haecceitism and essentialism help to clarify claims at issue in the debate between the complex and the simple view. But it is not the case that the simple view endorses, and the complex view denies, haecceitism and/or a minimal conception of persons' essential properties.

6.4 Conclusion

I have provided suggestions on how, if not in terms of the unhelpful labels 'complex' and 'simple', we should think about personal identity. My earlier suspicion was that a number of notions and distinctions are conflated under these rough labels. I now made this suspicion more precise. My examples include ontological dependence, fundamentality, essentialism, and haecceitism. As described, none of these concepts is free of difficulties or ambiguities. But all of them are better understood than the complex versus simple distinction. And as suggested by the exegetical evidence I gave along the way, all of these notions seem to play a role in the respective positions. In fact, reflecting on their details and variants, and the fact none of them neatly aligns with either the complex or the simple view, brought out how problematic it is to slide over these notions under the heading of one single distinction. My suggestions enable us to make the charge of conflation and confusion more precise.

My claim is not that understanding positions in terms of ontological dependence, fundamentality, essentialism, and haecceitism resolves ques-

tions or reservations we might have about their content. Even if these notions do not capture everything that is at issue in the cloudy complex versus simple distinction, they allow us to focus on specific differences between the positions that become apparent when being stated in these ways. My hope is that this leads to improvements in understanding, precision, and accessibility of the positions and their relations to each other. In this way, the notions could be seen as reference points for clearer and less ambiguous characterizations.

One advantage of replacing the complex versus simple distinction with the suggested notions is that it helps us to identify misguided commitments. For example, thinking that personal identity is fundamental might have led some to endorse the simple view. But as demonstrated, fundamentality is compatible with paradigmatically complex positions, and not an entailment of simple positions. Moreover, thinking that personal identity depends on other relations might have led some to call their position complex. But as demonstrated, paradigmatically simple positions too can understand personal identity as ontologically dependent. In other words, partly because of the opacity of ‘complex’ and ‘simple’, the views have sometimes been endorsed on grounds that properly understood must be deemed mistaken.

Given these insights and rephrasals, we can defend the simple view against charges of elusiveness and unintelligibility. Simple positions are not charitably understood as being dubious or vague, but as endorsing pictures that feature well-established metaphysical notions and their application to personal identity. It is a further question whether these pictures are true, and how could find out about this. What matters is that they

are not defective besides potentially being false, a feature which they share with their competitors.

Finally, let me return to future-directed concern. The suggestion to understand positions in terms of more precise notions can be fruitful towards evaluating the extreme claim according to which the complex view fails to give reasons for future-directed concern. One of the lessons from reflections on the complex versus simple distinction (chapter 5) was that given its confusions, it is unlikely to mark a difference in reason-giving power. It remains unclear which features of the complex view are accountable for the shortcomings in justifying special concern. In contrast, we are in a better position to evaluate the reasonableness of concern if personal identity is grounded in other relations, derivative rather than fundamental, and/or describable in purely qualitative terms. The contents of these versions are more precise than the extreme claim because reference to the complex view is replaced by better defined notions. We can thus discuss the extreme claim more fruitfully. However, while these versions of the extreme claim are easier to evaluate, our judgment should remain critical: it is not obvious at all how groundedness, derivativeness, or qualitiveness of personal identity should impact its importance.

Chapter 7

Personal Identity And Transformative Experiences

So far, I have been primarily concerned with the *numerical* identity of persons over time. But there are important kinds of *qualitative* changes which we all undergo in the course of our lives. They do not put an end to our existence or make us two rather than one and the same. Yet, they deeply change who we are, in the sense that they affect how we see the world and what matters to us. Paul describes the example of becoming a parent:

“[h]aving a child often results in the transformative experience of gestating, producing, and becoming attached to your own child. At least in the ordinary case, if you are a woman who has a child, you go through a distinctive and unique experience when growing, carrying and giving birth to the child, and in the process you form a particular, distinctive and unique attachment to the actual newborn you produce. Men can go through a partly similar experience, one without the physical part of gestating and giving birth. For both parents, in the usual case, the attachment is then deepened and developed as they raise their child. [...] When a newborn is produced, both parents experience dramatic hormonal changes and enter other new physiological states, all of which help to create the physical realizer for the intensely emotional phenomenology associated with the birth. These experiences contribute to the forming and strengthening of the attachment relation [...]. All of this generates the unique experience associated with having ones first child. Raising a child is then a temporally extended process that extends, deepens, and complicates this relationship. This unique type of experience often transforms people in the personal sense, and in the process, changes ones preferences.”⁴⁹⁶

Paul calls these *transformative* experiences. Because of the far-reaching changes they bring about in us, it is difficult if not impossible to know antecedently what it would be like to live the post-transformative life. This raises the question how we can ever rationally choose to undergo a transformative experience and become our post-transformative self. Paul’s answer is the revelation approach: choose based on how much

⁴⁹⁶Paul 2015, pp. 764-5.

value you assign to the revelation of making the novel experience.

In the following, I will briefly summarize Paul's discussion and the revelation approach, before turning to objections. Shupe is convinced that considerations about the numerical identity and distinctness of persons play a crucial role for the reasonableness of transformative choice: she grants that Paul's revelation approach can make sense of transformative choices for ourselves. But Shupe argues that when decision-maker and transforming individual are *distinct*, e.g., when parents consider an option that will be transformative for their child, the revelation approach is inappropriate. I will provide five strategies for responding to Shupe's objections. Moreover, I will show how Shupe's worries rest on a particular reading of the approach. I offer a more charitable interpretation of what Paul's approach demands.

7.1 Transformative Experience And Revelation

Paul argues that standard decision theory is in trouble.⁴⁹⁷ The theory holds that a rational decision-maker attempts to maximize expected utility, i.e. the product of the outcome's probability, given the decision-maker's action, and the utility it provides to the decision-maker. Paul objects that for an important class of decision problems, assessing and comparing expected utility is impossible.

Paul phrases her discussion in terms of a distinctive notion of value

⁴⁹⁷My overview of Paul's account will be rather brief. For more complete summaries, I recommend *Ibid.*; Barnes 2015, section 2; Pettigrew 2016; Shupe 2016, sections 2-6; Talbott 2016.

that shall be maximized: *subjective values* measure the value of an experience for an individual. They are

“grounded by more than merely qualitative or sensory characters, as they may also arise from nonsensory phenomenological features of experiences, especially rich, developed experiences that embed a range of mental states, including beliefs, emotions, and desires.”⁴⁹⁸

What it’s like to have a certain experience is an important (although not the only) determinant of subjective value. Two kinds of experiences raise challenges for the assessment of subjective value.

First, *epistemically transformative* experiences are such that prior to having them, one cannot know what they are like.

“When a person has a new and different kind of experience, a kind of experience that teaches her something she could not have learned without having that kind of experience, she has an *epistemic transformation*. Her knowledge of what something is like, and thus her subjective point of view, changes. With this new experience, she gains new abilities to cognitively entertain certain contents, she learns to understand things in a new way, and she may even gain new information.”⁴⁹⁹

As a result, if you never had an experience of this kind, you cannot determine the subjective value it would provide.

Secondly, a *personally transformative* experience changes the experimenter.

“If an experience changes you enough to substantially change your point of view, thus substantially revising your core preferences or revising how you experience being yourself, it is a personally transformative experience. [...] [I]f a personally transformative experience is a radically new experience for you, it means that important features of your future self, the self that results from

⁴⁹⁸Paul 2014, p. 12.

⁴⁹⁹*Ibid.*, pp. 10-11.

the personal transformation, are epistemically inaccessible to your current, inexperienced self.”⁵⁰⁰

Not only are you unable to assign subjective value to a transformative experience; sometimes having the experience *changes* the preferences. Any values one ascribes before the experience would *differ* (maybe even radically) from the values assigned *after* the experience.⁵⁰¹ One does not know in advance *how* they will differ.⁵⁰²

We could just rely on statistical data about how individuals tend to respond to certain experiences, and assume that we will respond in similar ways. Paul dismisses this suggestion on two grounds.

First, data is not informative enough. The shortcoming is related to the *reference class problem*:⁵⁰³ psychological, sociological, and economical data pertains to the level of populations or groups.⁵⁰⁴ From this, we cannot read off

“the kind of fine-grained information about how a person who is just like us, a person with just our particular blend of personal

⁵⁰⁰[Ibid.](#), pp. 16-7.

⁵⁰¹[Ibid.](#), p. 32.

⁵⁰²Consider the implications for Brink’s claim that the gold standard of compensation is its intrapersonal instance because it is *automatic* (3.1.2). I rephrased his claim as: when incurring present burdens and efforts, intrapersonal benefit will compensate you with maximal reliability. You work hard while being justified in believing that it will be *you* who will reap the rewards. What if it was you who will reap the rewards, but a future *you* who *has changed* in significant respects? The prospect of transformation gives a sense in which intrapersonal compensation is *not* automatic: you cannot know in advance how you will respond, in particular whether what you formerly considered a benefit will compensate you for past burdens.

⁵⁰³Paul 2014, p. 127.

⁵⁰⁴This character of the data is necessitated by the *fundamental identification problem*: one cannot observe one and the same individual simultaneously in a treatment and a non-treatment state. If the individual is treated, the effects cannot be compared with the effects of non-treatment of the very same individual at the very same time. Such comparison would rely on the evaluation of a counterfactual which presupposes knowledge about other, epistemically inaccessible possible worlds. The solution is to formulate research questions at the level of groups rather than individuals; cf. Paul and Healy [forthcoming](#), section 2.

abilities and personality traits, our likes and dislikes, our work ethic and neuroses, and so forth, is most likely to respond to a particular experience.”⁵⁰⁵

That is, such data does not tell you how *you* in particular will respond.

Second, even if data could provide such information, it would do so in an unsuitable way. Decision-making about important life choices should be *authentic*. Paul’s sense of authenticity requires that the decision-maker imagines each decision option and its consequences from her first-person perspective in order to determine whether she identifies with these outcomes and wants them to be a part of the continuation of her life story. Imaginative empathy is important for authentic self-realization.

“When you consider what might happen in your future, your consideration involves an imaginative reflection on what it will be like, from your point of view, to experience the series of future events that are the mostly likely outcomes of whatever it is that you choose to do. You use this reflection on what you think these events will be like, that is, what you think your lived experience will be like, to authentically determine your preferences about your future, and thus to decide how to rationally act in the present.”⁵⁰⁶

Considering third-personal data falls short of imaginative projection, and thus cannot ground authentic decision-making.

Expected utility theory presupposes that one can evaluate and compare different outcomes in order to maximize expected utilities. But post-transformative outcomes are inaccessible to us. Thus, no subjective values can be assigned, and no comparison can be made.

Paul thus argues that we have to rethink our decision theory. Suppose you are considering whether or not to undergo the epistemically

⁵⁰⁵Paul 2014, p. 132.

⁵⁰⁶*Ibid.*, pp. 106-7.

transformative experience of eating durian fruit for the first time. Paul proposes the following rephrasal of the decision problem:

“[t]he relevant outcomes, then, of the decision to have a durian *are discovering the taste of durian* versus *avoiding the discovery of the taste of durian*, and the values attached reflect the subjective value of making (or avoiding) this discovery, not whether the experience is enjoyable or unpleasant.”⁵⁰⁷

That is, Paul suggests to address the reformulated decision-problem by means of the *revelation approach*:

“you decide to try a durian for the sake of the revelation the experience of tasting a new kind of fruit brings.”⁵⁰⁸

Transformative choices can be rational if based on how one values the *discovery*, independently of whether or not it leads to pleasant experiences.⁵⁰⁹ If this value exceeds the value of leaving things as they are, it can be rational to make the transformative choice, independently of what it will be like.

“[Y]ou decide whether you want to discover how your life will unfold given the new type of experience. If you choose to undergo a transformative experience and its outcomes, you choose the experience for the sake of discovery itself, even if the experience, like having a child or enduring some sort of extended privation, involves stress, suffering, or pain.”⁵¹⁰

The best available way to make important transformative life choices is *not* based on projective imagination of what the outcomes are like, but on how one values the novelty of the experience independently of what it will be like. This motivates

⁵⁰⁷[Ibid.](#), p. 113.

⁵⁰⁸[Ibid.](#)

⁵⁰⁹[Ibid.](#), pp. 38-39, 114.

⁵¹⁰[Ibid.](#), p. 120.

“epistemic humility: life is more about discovery, and coming to terms with who we’ve made ourselves into via our choices, than about carefully executing a plan for self-realization.”⁵¹¹

7.2 Paul and Shupe On Cochlear Implants

Transformative choices do not only arise for our own lives. Sometimes we need to make life-changing decisions for others. Paul considers parents of a deaf child who are faced with the choice whether to provide their child with a cochlear implant.⁵¹² It is assumed that the child has been deaf from birth, and that the decision must be made within the first year of its life, given that the implant won’t be as effective later. Implantation can be irreversible, and may reduce or even eliminate any remaining capacity to hear naturally.

If all goes well, the child will get access to sounds and language. But there are also reasons against choosing the implant. The child would miss out on being a member of the Deaf community with its distinctive social structure—a culture that would be obliterated if everyone chooses the implant. Neither would the child enjoy the unique and rich experience that come with being deaf. Implants are getting better, but the hearing they provide is still not at the level of those who were never deaf. There is a risk that the child would feel or be perceived as “second-class, ‘disabled’, or once-disabled”⁵¹³. If the parents are themselves deaf, they might reasonably prefer to forego the implant to foster a deep and intimate parent-child relation.

⁵¹¹Paul 2015, p. 765.

⁵¹²Paul 2014, pp. 57-70.

⁵¹³*Ibid.*, p. 59.

This looks like an impasse about which option is most beneficial. But as Paul points out, it is even more tricky. One unrecognized difficulty is the transformative character of the decision problem. Hearing parents cannot know what it would be like to remain deaf and a member of the Deaf community. Deaf parents of a deaf child cannot know what it would be like to become able to hear. Neither do they know how an implant would shape their child's core preferences. They cannot evaluate the child's lived experiences in the outcomes. One reason is that different sense impressions at a time combine into one experience in a holistic way, and over time shape the individual's cognitive architecture. It is not the case that hearing individuals can just abstract away their hearing ability, and thereby arrive at what it's like to be deaf.⁵¹⁴ Just like in the vampire, parenthood, and durian cases, the normative standard for rational decision-making on the cochlear implant is unattainable.⁵¹⁵

On the one hand, this should make us doubt that “the parents can be expected to rationally evaluate these arguments”⁵¹⁶. On the other hand, the revelation approach has been introduced as a normative standard for “[w]hatever transformative decision we are considering”⁵¹⁷.

Shupe points at a difference between Paul's paradigm cases of transformative experience and the cochlear implant case. In the latter,

“the decision one is called to make is one that primarily affects the experiences and shapes the preferences of another person, not (just) oneself.”⁵¹⁸

⁵¹⁴[Ibid.](#), pp. 68-9.

⁵¹⁵[Ibid.](#), p. 84.

⁵¹⁶[Ibid.](#), p. 62.

⁵¹⁷[Ibid.](#), p. 120.

⁵¹⁸Shupe 2016, p. 3124.

The decision-maker behind the transformative choice is not identical to the individual affected by the choice. Paul is not discussing this structural dissimilarity among the cases.

Shupe thinks that it matters. Paul presents the revelation approach as an “all-purpose strategy”⁵¹⁹ for making transformative choices. But according to Shupe, the cochlear implant case shows that this understanding is false. When decision-maker and transforming individual are distinct, the revelation approach is implausible. She writes:

“in order for the revelation approach to be appropriate, it would have to be the case that discoveries occasioned by the agent’s choice really were the kinds of thing the agent might legitimately value independently of her choice’s consequences. And, in the cochlear implant case, I do not take this to be plausibly the case. [...] [C]onsider the possibility that the parents do decide to give their child the cochlear implant and she grows up to be deeply unhappy as a result of their choice. Would it be rational for the parents to discount these possible harms when making their decision, and to make their choice only on the basis of the revelatory value of experiencing the consequences of their decision and the value of discovering how their own preferences will evolve as a result of their choice?”⁵²⁰

I read two worries from this quotation. First, it is possible that the child receives the implant⁵²¹ and grows up unhappily. It would seem odd to ignore such consequences and choose on the basis of revelation. Shupe grants that agents can legitimately choose on the basis of revelation, i.e. for the sake of discovery itself, and screen off consequences—e.g., privation, stress, pain—when making transformative choices *for themselves*.

⁵¹⁹Shupe 2016, p. 3132.

⁵²⁰*Ibid.*, p. 3125.

⁵²¹For the sake of simplicity, I will refer to the example of hearing parents choosing a cochlear implant for their deaf child. However, Paul, Shupe and I agree that the same difficulties arise in cases where deaf parents choose such an implant, or where either deaf or hearing parents choose to forgo an implant for their child.

But as quoted, she thinks this is different when making such choices for others. I call this the *consequences* objection.

Secondly, a literal application of the revelation approach would require the parents to make the decision in view of the revelation which choosing the implant for their child would provide to *them*. This is an odd demand. It is a significant change to have one's child turn from deaf to hearing, but the much more important transformation—intuitively and according to Paul herself⁵²²—is experienced by the child. I call this the *parent-focus* objection.

Shupe describes and dismisses two alternative applications of the revelation approach. Both suggest a form of *surrogate decision-making* where the parents would assess the revelatory values on behalf of the child. However, she argues that the readings face serious obstacles. If one attempts a *substitute judgement* on the basis of the child's preferences and values, the problem arises that

“the infant does not yet have a preference for either of the revelatory outcomes in question; furthermore, [...] the infant's future preferences [...] will largely be shaped by what her parents decide to do.”⁵²³

Alternatively, if one attempts to focus on the child's *best interests*, the problem arises that it is very difficult to predict which option is best for the child.

Shupe⁵²⁴ shows that in real life, many transformative experiences are more like the cochlear implant rather than the parenthood case. Doctors make transformative choices on behalf of patients who lack decision-

⁵²²Paul 2014, pp. 59-60.

⁵²³Shupe 2016, p. 3126.

⁵²⁴*Ibid.*, sect. 9.

making capacities, for example when they weigh treatment options with different risk-benefit ratios and one or more possible transformative outcomes, such as death or disability. Parents make transformative choices on behalf of children, for example in interfaith marriages when they decide about the faith in which their child shall be raised. Leaders and representatives make transformative decisions on behalf of their fellow group members, for example when they decide about whether or not to participate in an uprising. Thus, if Shupe's criticism is right, there are many situations in which the revelation approach seems insufficient to guide our decisions. Shupe demands that Paul should be clearer about the limitations of the approach.

Before I comment on Shupe's objections, I note how they are in tension with some claims I defended so far. Personal identity and distinctness are less relevant to practical concerns than many think. Distinctness amongst persons and their stages does not preclude justified concern (chapter 2). And identity by itself is not enough unless we explain how it relates to practically important relations (chapter 3). Shupe now gives a sense in which personal identity *is* central to the reasonableness of transformative choices. When making them by means of the revelation approach, it matters that the decision-maker and the individual undergoing the transformation and experiencing the revelation are one and the same person. If not, then the revelation approach is unsuitable. In other words, the right way to make transformative choices for myself is deeply inadequate for making transformative choices for someone else. First, this suggests that for transformative choices, the key claim of this dissertation is false, and personal identity is central to their rationality.

Secondly, transformative choices for others remain rationally intractable.

However, most of Shupe's criticism can be resisted. I shall now discuss her *parent-focus* and *consequences* objections, before addressing difficulties of transformative surrogate decision-making. In the process, I will suggest five strategies for responding to Shupe.

7.3 Parent-Focus

Shupe thinks the revelation approach asks the parents to choose the option that provides most value *to them*. According to the *parent-focus* objection, this is inadequate. I will now sketch two strategies for arguing that even if the revelation approach is parent-focused in this way, this is no reason to reject it.

Both strategies share the following starting point. Campbell objects that the kind of imagination which Paul thinks is distinctive of authentic decision-making overstates the importance of the decision-maker's future lived experiences. One of Campbell's examples is a high-school teacher who plans to work in "a bad part of town."⁵²⁵ Her own anticipated subjective experiences, the teacher says, are *irrelevant* to her authentic choice. What matters to her is that the kids are better off. That is, we can make perfectly authentic choices *without* considering our lived experiences.

In response, Paul clarifies that we should not understand subjective values too narrowly. They can reflect more than our own lived experiences, for example the experiences of others. For example,

"[i]f I make a medical decision for my child, my decision is heavily

⁵²⁵Campbell 2015, p. 791.

influenced by my assessment of the quality of her future lived experience”⁵²⁶.

The parents’ subjective assessment of the options can be a function of how they judge the quality of their child’s respective lived experiences. Typically, if a child enjoys valuable lived experiences, the subjective values the parents enjoy will be increased.

7.3.1 Strategy 1: Parent-Focus Is Not Implausible

This clarification allows us to see that the *parent-focus* objection need not be problematic. When the parents evaluate options like those in the cochlear implant case, it is not like the child’s lived experiences do not matter or should come second. The approach is perfectly compatible with a dependence of the parents’ assessment on the child enjoying valuable lived experiences. The approach is also perfectly compatible with norms prescribing that the parents shall take the child’s lived experiences into account when making the decision, or make the child’s experiences the sole ground of their choice. In other words, even if we frame the decision about the implant as one that focuses on the parents’ revelatory values, we need not exclude or mitigate the importance of the child in the parents’ deliberation.

Paul’s novel diagnosis is that the case is complicated by the decision-makers’ ignorance of what at least one of the options is like. For this reason, the parents face deep obstacles when evaluating the child’s lived experiences. One might worry that this stands in the way of a proper, reliable dependence of the parental decision-making on how the child will

⁵²⁶Paul 2015, p. 810.

be doing. This is Paul's worry. The parents are radically ignorant of the facts that should influence their judgement. A sound decision-making process would proceed "by *relying on their own subjective evaluation of the situation*"⁵²⁷ and "by projecting forward into the imagined future lived experience of the child under each possibility"⁵²⁸. For at least one of the options, this is impossible.

But it is another question whether this leads to an implausible parent-focus or -bias. As just stated, Paul agrees that the parents' decision is, and should be, influenced by their assessment of the child's lived experiences. And we might just deny that the parents' shortcomings in projective imagination preclude that their assessments track the child's lived experiences. Maybe something *less* than projective imagination *suffices* to track the child's lived experiences well enough to make a rational decision in which parent-focus and concern for the child's subjective values go hand in hand.

According to one plausible understanding of this influence, the parents' revelatory values are a function of how their child is doing. For example, we saw how Brink compares the relation between parent and child to the relation between fissioner and fission products (3.1.2). He agrees that fissioner and fission products are distinct, but because they share experiences, beliefs, values, desires and actions, they are justified in seeing their individual interest as extended across the person boundaries, by the interests of the other.⁵²⁹ Through such extensions, even rational egoists who focus primarily on their own good—which I suggest

⁵²⁷Paul 2014, p. 67.

⁵²⁸*Ibid.*, p. 65.

⁵²⁹Brink 1997b, p. 126; *ibid.*, p. 141.

as the analogue to parents who focus on their own revelatory values—can come to have a “derivative but non-instrumental reason to be concerned about others.”⁵³⁰ What Brink means is that an agent’s consideration of the good of other, psychologically continuous beings is not just a useful means, but valuable for its own sake in virtue of being constitutive of the agent’s own good.⁵³¹

This provides one way to see parental values as dependent on the lived experiences of the child. It need not require that the parents know exactly what the child’s future lived experiences are like, as if they themselves once had a similar experience. In order to secure the needed influence, it would suffice if they know *that* their child’s future experiences are valuable. Let me give an analogy. If I am happy only if you are happy, then for evaluating a possible future, it suffices that I know *that* you are happy; I do not need to know *what it is like* for you to be happy. In the implant case, too, if my subjective values depend on the value of the child’s lived experiences, it is plausible that I do not need to know the exact nature of these lived experiences from the inside or through projective imagination to determine whether or not a given possible future is subjectively valuable to me. For the parents to evaluate a given option, it would suffice to know *that* the child will be contented if the option is chosen.

This suggestion is not intended to address or remove uncertainties about how the child will be doing in a given scenario. Shupe mentions three issues that seem almost insurmountable.⁵³² For example, one could

⁵³⁰Brink 1997a, p. 127.

⁵³¹Brink 1997b, p. 147.

⁵³²In light of these difficulties, Shupe argues that postponing the decision until the

attempt to compare the likelihood of a happy life with an implant versus a happy life without an implant.⁵³³ Unfortunately, these likelihoods are difficult to determine. Moreover, satisfaction and happiness might be incommensurable across the two scenarios. And even if they were comparable, it is unclear by which measure of happiness they should be compared. As a result,

“the parent cannot reliably discern the appropriate attitude towards revelation to take up on behalf of their infant”⁵³⁴.

I agree with these worries. But I deny that they are unique to transformative choice and the revelation approach. Quite similar difficulties complicate my deliberation about whether to go to the cinema tonight or rather read a book at home, and the subsequent decision to pursue one of the options on the basis of maximizing my subjective rather than revelatory value. Transformative choices can be incredibly difficult, but if anything makes them more difficult than non-transformative choices, and if anything makes the revelation approach less eligible to address them, certain other characteristics than these uncertainties would have to be responsible. The revelation approach’s parent-focus is neither implausible *per se* nor makes it more susceptible to difficulties related to uncertainty.

child can make it herself looks like the most reasonable option, but unfortunately it is unavailable for time-sensitive interventions.

⁵³³See also Pettigrew 2016.

⁵³⁴Shupe 2016, p. 3127.

7.3.2 Strategy 2: The Revelation Approach Does Not Apply

Shupe notes that there might be some sense in which the parents would make a discovery when providing their child with the implant. But

“these are not the epistemically inaccessible features of the choice that seem most important to us.”⁵³⁵

The salient and important change is experienced by the child.

This thought, I argue, can be pushed one step further: it is not obvious that the parents will have a transformative experience in either scenario. If so, their decision problem is not one of *transformative* choice. This seems at odds with Paul’s characterization of the case, but let me explain why she could be sympathetic to this suggestion.

The child is doubtlessly facing a transformative experience. Receiving the implant and becoming able to hear would provide it with radically new sensory abilities. Foregoing the implant would be transformative, too. It will enable a distinctive cognitive development and membership in the Deaf community and thus provide experiences whose nature the child currently cannot imagine. So the child will definitely transform.

This is not obviously the case for the parents. Having their child turn from deaf to hearing certainly affects them profoundly. But is it a change that provides them with radically new epistemic perspectives or changed core preferences? I think it need not. Although there will be change, we need not count it as transformative. Of course, we said that they cannot know what it’s like to receive a cochlear implant, and

⁵³⁵Shupe 2016, p. 3125.

so their representation of their *child's* lived experiences in the possible futures will be radically incomplete. But according to Shupe's reading, the revelation approach asks them to focus on *their own* lived experiences anyways, not the ones of the child.

Here is one way to make sense of the suggestion that parents' experiences will not be transformative. If they were facing a transformative choice, they would be unable to assign subjective values to at least one of the outcomes. And I claim that we can defensibly deny this for the cochlear implant case. Suppose that, as suggested earlier, their own subjective values in the outcomes depend on the lived experiences of the child. Then even if the parents cannot projectively imagine what the child's experiences will be like, their subjective values could still be perfectly determinate. If I want to make you happy, and have exactly two options, both of which I know make you happy, then I can perfectly well assess my subjective values while being completely ignorant of what it's like for you to be happy. Similarly, if the parents had third-personal data on how likely the child will be contented in either scenario, it seems their subjective values could be determinate. If the parents can assign subjective values, why should they refer to the revelation approach?

Shupe points out that these probabilities are hard to get. However, we should expect this to be everybody's problem, not one that is unique to transformative experience or the revelation approach. It wouldn't be a reason to think the revelation approach fares worse than other approaches.

Again, the suggestion that the parents' decision is not transformative *for them* does not make their task easier. They remain confronted with

risks and uncertainties about whether or not their child will be contented. And the child's transformative experience certainly makes it harder to predict how the child will respond. But the present suggestion locates the difficulty for the parents' decision problem not in any transformative character of *their* decision-problem, but in the fact that it is quite unclear which option will be better for the child.

Strategy 1 denies that the parent-focus is implausible. Strategy 2 argues that even if the revelation approach's parent-focus would be implausible, this is not a problem. The revelation approach is designed to deal with *transformative* choices, not choices *tout court*. Shupe's reading brings into view the possibility that the parents do not face a transformative choice. If so, they need not refer to the revelation approach in the first place.

In what follows, I will not insist on this suggestion, and proceed under the assumption that the parents do in fact face a transformative choice of some sort, and that they are thus in need of the revelation approach.

7.4 Harms And Other Consequences

Shupe helpfully points out that not all transformative experiences are the same. If I eat durian for the first time, only I am affected. If I decide to become a parent, not only do I change fundamentally; I also affect another life in the most fundamental way: by bringing it into existence. Choosing the cochlear implant might affect the parents, but it is a much more significant change for the child. According to Shupe's *consequences* objection, the revelation approach requires that these effects shall be

ignored or discounted by the decision-maker.

I see two strategies for resisting the objection. Again, both have a common starting point. Barnes writes that some decision problems contain options where the decision-maker does not know what it's like to experience the outcome, yet knows *enough* about this option to determine that she does not want to experience its outcome. The decision-maker can arrive at this conclusion *without* thinking about revelation.

“When I choose not to have children, it isn't simply a choice to avoid 'revelation' or a choice made 'for the sake [or lack of] discovery itself'. It is, quite simply, a choice not to have children because I know I don't want them. And so it's a choice based on projected outcomes. I can project, given my current desires and presences [*sic*], that having kids is a less good outcome than not having them, even though I'm ignorant of some of the relevant subjective values.”⁵³⁶

Paul accepts that we are entitled to avoid experiences like getting eaten by a shark even if we cannot know antecedently what it is like to have them,⁵³⁷ but she denies that this shows that the revelation approach is not at work.⁵³⁸

7.4.1 Strategy 3: Enrich Revelatory Values

Some revelatory values are assigned to transformative options independently of what the resulting experiences will be like. In particular, they can ignore whether or not the resulting experiences will be pleasant or not—information which due to the transformative character of the experience is not accessible. Instead, some revelatory values turn solely

⁵³⁶Barnes 2015, p. 785.

⁵³⁷Paul 2014, pp. 27-8.

⁵³⁸Paul 2015, p. 805.

on how much the decision-maker values discovering the novel kind of experience.

To some degree, the value which a decision-maker assigns to a given novel experience can certainly reflect *expectations* about how one will respond. Barnes and Paul agree that for some transformative experiences, although the decision-maker cannot know what it will be like to undergo them and how the experience will change her, the decision-maker knows *enough*. In contrast to the vampire, parenthood, and durian cases, when contemplating to swim with sharks, it is clear that the—in Barnes' terminology—'projected outcome' will not be a good one. Even without projectively imagining the experience and its consequences, avoiding it is better than choosing to undergo it.

But Paul shows how this is compatible with the revelation approach. Expectations of this kind can feed into revelatory values which we assign to making the novel experience. In other words, Barnes' 'projected outcomes', which need not involve the first-personal imagination of experiencing the outcome, can factor into how much we value the novelty. Because swimming with sharks is likely to be unpleasant, the revelatory value of doing it is low, and exceeded by the value of the status quo.

Besides these somewhat inarticulate but clear enough subjective features of experiences like the shark encounter, we may be inclined to avoid other transformative experience because they conflict with our values. For example, Winston in Orwell's *1984* takes an interest in avoiding getting brainwashed by the Thought Police⁵³⁹ not only because it may be unpleasant, but because it conflicts with the way he and us value

⁵³⁹Barnes 2015, p. 781.

autonomy and self-determination.⁵⁴⁰

For the cochlear implant case, this means that the *consequence* objection against the revelation approach can be addressed. Expectations about harms and other consequences can feed into how much the parents' value the revelatory decision options. If the parents were perfectly clear that choosing the implant has bad 'projected outcomes', then this can affect their revelatory values.

A second point to mention is that the approach does not make the substantive requirement that parental discovery matters more than the consequences for the child. Just like standard decision theory, there is a sense in which the revelation approach is *agnostic* on how the parents shall rationally prioritize between their own discoveries and the well-being of the child. It depends on their preferences, values, and expectations. Of course, these can be constrained by moral, epistemic, and societal norms. The point is that the approach by itself does not require the parents to discount possible harms for the child in view of their own discoveries.

7.4.2 Strategy 4: The Revelation Approach Does Not Apply

A slightly different strategy to address the *consequences* objection according to which the revelation approach implausibly requires the parents to discount harms to their child is the following. Recall that the revelation approach covers decisions in which rational and authentic decision-making is impossible because the lived experiences in at least one of

⁵⁴⁰Paul 2015, pp. 803-4.

the outcomes are inaccessible. As outlined, Barnes and Paul agree that sometimes, we know *enough* about the features of a transformative experience to determine our revelatory values and to make a decision on this basis. But alternatively, we could come to think that sometimes, we know enough about a transformative experience that we are not in need of the revelation approach.

One way to understand Shupe is that revelation is entirely inadequate to evaluate cochlear implants. We want the child to lead a happier life, not just to enjoy the more intriguing revelation. Here, one factor might be that the child will be unable to experience its development with or without the implant *as a* revelation. Not only might the very young child be *unaware* of being in the process of having either auditory stimuli or inclusion into the Deaf community revealed to it. The child is also necessarily unfamiliar with the respective alternative, given our assumption that either revelatory option requires a decision at the early stages of the child's development *and* will be irreversible. These difficulties might support the intuition that not revelatory values, but evidence about which option would provide the child with superior—e.g., happier—lived experiences would give us what matters. In that case, the benefits of that option and the corresponding harms in the non-ideal scenario should not be discounted in view of revelation.

However, I suggest that if we know which option leads to superior lived experiences, maybe we should not be thinking about revelation in the first place. We would be able to identify the option we want without thinking about revelatory values, which can indeed seem rather unimportant in comparison. This kind of reasoning seems to be behind Barnes'

suggestion that Winston is not thinking about discovery or revelation when avoiding the Thought Police: he knows that getting brainwashed will have worse consequences for him. Because of this, he has no need to think about revelation. According to this reading, discounting consequences does not bring the revelation approach under pressure, because the revelation approach is silent on how we should choose if we do know the consequences of each option in sufficient detail.

Paul would probably not take this stance. Instead, she would put forward Strategy 3, i.e. apply the revelation approach, but highlight how rich the notion of a revelatory value can be.⁵⁴¹ After all, the revelation approach is intended to apply to transformative choice in general, not just to a subclass of transformative choices where it is unknown which option has better consequences. Nevertheless, I think it would not be a costly concession to accept Barnes' idea that the case for or against *some* transformative options is so clear that we do not need the revelation approach in these instances. The approach would still apply in many cases, especially the ones that are not straightforward and where considerations about consequences do not help.

The cochlear implant case is of the latter kind. What makes it so difficult is the lack of straightforward 'projected outcomes', current desires and preferences, or other expectations about the future that clearly favour one of the options and could inform our choice. This means that proceeding without considering these aspects further is not a shortcoming. The intractability of the 'projected outcomes' is what makes the

⁵⁴¹Then again, Paul also assumes "that cases like the shark-eating case are outside the scope of [her] discussion" (2014, p. 32). Maybe if the consequences of the transformative experience are obvious, she can live with bracketing the revelation approach.

revelation approach necessary in the first place. If we had a clue about the consequences, we would be entitled to proceed without considering revelation. In other words, whenever the problem described by Shupe arises, the revelation approach does not apply.

Given the availability of these two strategies, i.e. building expectations about consequences into revelatory values, and denying that the revelation approach applies if expectations about consequences clearly favour one of the options, we can conclude that there are ways to retain the revelation approach without ignoring or discounting harms and consequences.

7.5 Surrogate Decision-Making

Let us assume that the strategies considered so far are unsuccessful, and that the difficulties point at a different interpretation according to which the parents should attempt some form of surrogate decision-making for the child. Shupe thinks that such judgments are impossible in the cochlear implant case: the parents cannot apply the revelation approach for the child because the child does not yet have preferences about the outcomes, and her future preferences will be shaped by what the parents choose.

I will later disagree with Shupe's assessment that the revelation approach cannot be applied in the cochlear implant case (7.6). But first, I explain why Shupe is correct in pointing at challenges of surrogate decision-making in these contexts.

In order to decide on medical treatment for patients who are incompe-

tent, i.e. unable to decide for themselves, legal and ethical codes require doctors and researchers to seek the judgement of a surrogate decision-maker. The validity and moral authority of the surrogate judgement is typically thought to derive from satisfaction of one of the following standards. First, the surrogate decision-maker is in a good position to judge what is in the patient's best interest, and shall decide accordingly. Secondly, the surrogate decision-maker is the best source of knowledge about the preferences of the patient while she was competent, and shall decide according to these preferences. I will discuss these standards in turn.

7.5.1 Best Interests

The best interest standard requires the surrogate decision-maker to choose the option that provides the highest net benefit towards the patient's quality of life.⁵⁴² The surrogate shall decide on behalf of the patient, and, in contrast to the substituted judgement standard, not necessarily in the way the patient would decide herself.

The notion of quality of life is not without difficulties.⁵⁴³ First of all, due to the patient's incompetence, the assessment of her quality of life is external and might well deviate from the patient's subjective well-being. Secondly, quality of life has more than one dimension, e.g., freedom from pain, independence, mobility. It is not obvious how to weigh these dimensions against each other. Third, there is more than one standard of what constitutes quality of life: the pleasantness of the

⁵⁴²Beauchamp and Childress 2013, p. 228.

⁵⁴³Brock 1993.

individual's experiences, the satisfaction of her preferences,⁵⁴⁴ or a list of objective goods.

Presumably, what well-being means for children differs slightly from what it means for adults:

“children's well-being depends less on their current individual preferences and more on the objective conditions necessary to foster their development and opportunities than does the well-being of adults.”⁵⁴⁵

Buchanan and Brock argue that childrens' manifest and rather transient goals and preferences should not be taken as determinants of their well-being, given that children lack aims, values, and commitments. Instead, fostering their well-being means enabling them to develop capacities for reflective and independent judgements that allow them to exercise self-determination, to develop and implement life plans, and to choose and pursue aims and values.

There might be a tension between the best interest standard and the revelation approach. The latter asks us to consider the options independently of how their consequences feel and affect our happiness and well-being, whereas these things are what matters according to the former. However, Strategy 3 can resolve this tension: expectations about whether a revelation is in our best interest can factor into our assessments of the revelatory values.

Unfortunately, it is not clear what this means for the cochlear implant case. We cannot tell how the revelatory options will impact the

⁵⁴⁴If her preferences are not known, we can appeal to “what most informed and reasonable people would choose in the circumstances” (Brock 1993, p. 112).

⁵⁴⁵Buchanan and Brock 1990, p. 228.

child's well-being. Shupe cites Savulescu⁵⁴⁶ who thinks that providing the implant is almost always in the child's best interest. I will not evaluate his arguments; instead, I note that in light of the disagreement and controversies described by Paul, it seems hard to deem one of the options clearly superior. As it stands, we might have to suspend judgment. If so, the best interest standard is not very helpful for advancing our surrogate decision-making in the cochlear implant case.

7.5.2 Substitute Judgement

According to a different standard, proxy consent is valid if it is a decision the competent patient would have made.⁵⁴⁷ The proxy is expected to evaluate the decision from the perspective of the patient when she was healthy and competent. In the ideal case, the proxy judgment is the best approximation available of what mattered to the patient. However, it can be challenging to evaluate whether a proxy judgement achieves this goal. For example, a proxy might

“selectively choose from the patient's life history those values that accord with the surrogate's own values, and then use only those selected values in reaching decisions. The surrogate might also base his or her findings on values of the patient that are only distantly relevant to the immediate decision.”⁵⁴⁸

There is empirical evidence that surrogate decision-makers are indeed much worse than we would hope at anticipating the preferences of the individuals who nominated them.⁵⁴⁹ Especially when the patient did not

⁵⁴⁶Savulescu 2009.

⁵⁴⁷See for example Brock 1999, p. 521.

⁵⁴⁸Beauchamp and Childress 2013, p. 228.

⁵⁴⁹Seckler et al. 1991.

previously articulate her relevant preferences or values, the substituted judgment standard runs the risk of being “highly reconstructive, even speculative”⁵⁵⁰, and “contrived”⁵⁵¹.

One difficulty is the following. A substitute judgement to do X is valid if

in the closest possible world where the patient is able to offer consent, she would choose X *in the actual world*.⁵⁵²

Unfortunately, this counterfactual is difficult to evaluate, especially if the patient has never articulated relevant preferences before becoming incompetent that could now be respected by the surrogate.

In the case of young children and other patients who have never been competent before, there is a severe extra complication. As Nagasawa argues, the counterfactual is not only difficult but *impossible* to evaluate.⁵⁵³ Firstly, we have no way to tell what the closest world would be like in which such a patient is competent, e.g., whether it would differ in its laws of nature from the actual world, or whether only certain contingent facts are different. Secondly, we have no evidence whatsoever from which to infer which judgement the patient would make in that possible world, given that in the actual world she has never been competent. Thirdly, closeness is determined by a similarity relation whose exact nature is a

⁵⁵⁰Buchanan and Brock 1990, p. 119.

⁵⁵¹Harris 2003, p. 11.

⁵⁵²The need to distinguish the world in which X is being carried out from the world in which the subject chooses is recognized by Barnbaum (1999, pp. 170-2), and Nagasawa (2008, p. 20). They react to Wierenga (1983) who argues that if counterfactual competence and the considered treatment are thought to obtain in the same world, the substitute judgement gives inadequate results. For example, the counterfactual ‘If the patient was competent, she would choose withdrawal of life support’ would almost never license withdrawal of life support. This is because almost no competent patient would want her life-support to be removed.

⁵⁵³Nagasawa 2008, pp. 21-2.

contextual matter. More than one world can count as closest in a given context. It might well be that two worlds are both closest to the actual world, yet *differ* with regards to the patient's judgement.⁵⁵⁴

Another way to put the point is provided by Buchanan and Brock. They argue that the purpose of substitute judgements is to exercise and extend the patient's capacity for self-determination. But if the individual was never competent and has never been in a position to live a self-determined life, to form and to deliberate on preferences, then it is a mistake to think that such a substitute judgment is sensible. There simply is nothing it could approximate or emulate.⁵⁵⁵

In principle, substitute judgments could be made about transformative experiences. The surrogate decision-maker would have to be in a position to imagine the revelatory values of the now incompetent patient. But in the cochlear implant case, the standard is unhelpful. It is hard if not impossible to determine the child's counterfactual judgements on the intervention, given that it has never been competent before.

7.5.3 Strategy 5: The Revelation Approach Is Innocent

It is unclear whether the revelation to receive a cochlear implant is in the child's best interests, and making a substitute judgement is not possible.

Does this put the revelation approach under pressure as a candidate

⁵⁵⁴See also Wrigley 2011, pp. 179-80, 182-3. His chief worry is that there are many contexts where the validity of the substitute judgement will be underdetermined, given that there are many plausible similarity relations for evaluating the respective counterfactual.

⁵⁵⁵Buchanan and Brock 1990, pp. 113-7. This worry is echoed by Beauchamp and Childress 2013, p. 227 and Wrigley 2011, p. 180.

criterion for making transformative decisions for others? I think not.

I see standards like the best interest and substitute judgement standards as potential *complements* to the revelation approach. While the substantive demand of the approach is to focus on revelation, the standards determine which revelations have the most desirable features and should thus be chosen. In the cochlear implant case, the standards are not too helpful towards answering this question. But we can hardly fault the revelation approach for failing to answer questions which the standards struggle to address independently of which approach they are complementing. Indeed, the standards are uninformative regardless of whether they evaluate revelations or something else. The fact that we sometimes encounter outcomes for which it is very difficult to assess their revelatory value does not establish that revelation is the wrong thing to focus on in transformative choice. We can admit that the cochlear implant case shows how the revelation approach is not an all-purpose solution in the sense that for each transformative decision-problem, it will resolve all our questions about what to do. But difficulties like the complications when applying the revelation approach in surrogate decision-making demonstrate that we are dealing with an incredibly difficult decision problem, *not* that the revelation approach is defective or fares worse than other approaches.

In short, the present strategy admits the difficulty, but maintains that it arises independently of the revelation approach. The latter does not struggle with the difficulty more than competing approaches.

7.6 A Proposal On Cochlear Implants

We saw that surrogate decision-making for individuals like young children who have never been competent is complicated, but that any decision-making approach faces these complications. I implicitly conceded that the revelation approach does not help us to make a surrogate judgement in the cochlear implant case. This might be too modest.

I discussed Shupe's parent-focused reading (7.3) according to which the approach requires the parents to reflect on how they would value having their child transform. But there is an ambiguity in the revelation approach. It asks the decision-maker to decide based on how she assesses the revelatory options. This allows for a different understanding of what is required. The revelation approach can be read as asking the parents to decide based on the revelatory values *of the child*. This suggestion is sensible. If, as the revelation approach suggests, the "value of some experiences comes from what they teach us,"⁵⁵⁶ then why should the value of my child's transformative experience not come from what it teaches my child?

Shupe's and my reading agree that the parents assess the options, and the child is the transforming individual. But the readings differ in whose utility matters. Shupe focuses on the parents, whereas I argue that the child's revelatory values are what matters. Thus, my present suggestion is not equivalent to Strategy 1, which makes the parental assessments a function of the child's values. Instead, the parents are asked to try to discern *the child's* values, and choose the best option on this basis.

⁵⁵⁶Paul 2014, p. 91.

As with any transformative experience, there is a sense in which the parents' judgement on this question will be poorly informed. The options will be epistemically inaccessible to them. If they are able to hear, they do not know what it is like to be a member of the deaf community. If they are deaf or hearing, they do not know what it's like to turn from deaf to hearing. However, this need not make decision-making impossible. It is precisely the suggestion of the revelation approach that, while inaccessibility should motivate epistemic humility, *something* can be said. In the parenthood, vampire, and durian cases, the decision-maker too cannot know what the revelation will be like, yet the approach assumes that judgements are possible on whether the decision-maker prefers experiencing the novel and unknown over leaving things as they are. This shows how the fact that the parents cannot know what one of the revelatory outcomes is like is not what makes the cochlear implant case special or harder than other transformative choices. According to the approach, the decision shall be made based on whether a particular novel experience is preferred independently of what it's like to live through the experience and its consequences.

What kind of surrogate decision-making is being suggested here? It is certainly not a substitute judgement, given that for patients who have never been competent, the notion of a substitute judgements is difficult to make sense of. But it is not a clear-cut case of a best interest assessment either. We assumed that the evidence on which option is best in terms of the child's quality of life is not decisive. In the proposed application of the revelation approach, the parents are thus not deciding on the basis of such considerations. It would arguably be ideal to choose what is in

the child's best interest, but it is hard to imagine which option this is. For this reason, the revelation approach recommends choosing one of the options for its own sake. It thus fits *neither* of the two common standards for surrogate decision-making. Given their unhelpfulness in the present case, this should be seen as a feature rather than a shortcoming of the revelation approach.

If this judgment is made by the parents, it will be made against the backdrop of their own views and values. For example, given that the parents are able to hear, presumably do not find hearing deeply dissatisfying, are not acquainted with what it's like to be a member of the Deaf community, and can be expected to value having a child with the same sensory abilities as them to foster intimate connections,⁵⁵⁷ they might lean towards choosing the implant. But note that this need not be the case. Other hearing parents might well be convinced that growing up as a member of the Deaf community is the better revelation for their child.

The approaches and standards fail to clearly favour one decision option over the other in the cochlear implant case. I think there are two ways to make sense of the resulting impasse. We could see it as posing a dilemma. Regardless of whether or not we choose the implant for the child, we deprive it of rich and invaluable experiences. But alternatively, we could see these insights as liberating. Neither option seems clearly better than the other in light of the child's interests and preferences. Thus, regardless how we choose, we are not foregoing a superior

⁵⁵⁷Paul 2014, p. 61.

option.⁵⁵⁸ I am not suggesting that it does not matter which option we choose, but that there are good motivations for either option, and that the decision is up to us in the sense that the considered approaches and standards do not clearly constrain our decision-making.

In either case, the question arises whether it is legitimate for the parents to make a transformative decision in the light of *their own* values when deciding about the transformative experience *of their child*. Here, I would like to point at a brief discussion by Buchanan and Brock.⁵⁵⁹ First, parents are the best guides to their child's welfare. They care about and know it as well as anyone. The parents are thus in a unique position to discern the child's needs. Secondly, parents bear some of the consequences of the treatment decision, including financial costs. It is thus appropriate to grant them some authority on which treatment the child shall receive. The child's interests and the importance of gaining abilities for self-determination are not the only things that matter. The *parents'* interest in what happens to their child need to be considered, too. Thirdly, within limits, parents arguably enjoy the right to raise children in accordance with their own values. Children lack values and decision-making abilities of their own, before acquiring them through socialization and development.

“Someone must inevitably shape children's goals and values, however, and since we assign child-rearing responsibilities in our society largely to the family, it seems reasonable to accord to the family as well some significant discretion in imparting its values

⁵⁵⁸Further considerations besides the particular child's interests and preferences, such as the intrinsic value of the Deaf community and a corresponding imperative to preserve it, might well move the needle towards one option rather than another.

⁵⁵⁹Buchanan and Brock 1990, pp. 232-4.

to the children within it.”⁵⁶⁰

I see at least the following implications for the cochlear implant case. Parents too bring important interests into the context of the decision-problem. It is thus not unreasonable to take these interests into account and consider how they would be affected by the child’s transformative experience, even if these are only some alongside further important interests and considerations. Moreover, the fact that within limits, parents may raise their child in accordance with their own values entitles them to project their own evaluations into the decision-problem when making a decision on behalf of the child.

I note that if we believe that children do not yet exercise self-determination, there are no values and decision-making capacities which the parental judgment would override. It is thus not the case that parental decision-making interferes with the deliberation of the individual whose treatment is under discussion.

According to the reasoning described by Buchanan and Brock, the child starts out as a “tabula rasa”⁵⁶¹, and through raising her, the parents are entitled to transmit their values to the child. I suspect that at some point, we might be concerned that the child has ceased to be a tabula rasa, and seems to adopt preferences and to subscribe to views and values. I suggest that at least as long as these fall short of constituting proper decision-making capacity and authority, we need not see parental surrogate decision-making as overriding or interfering. We can recognize apparent values of the child as causally continuous and connected

⁵⁶⁰*Ibid.*, p. 233.

⁵⁶¹*Ibid.*, p. 234.

with parental values, and this could license a presumption of overlap in core values and preferences. Parental decision-making could be seen as expressing and extending, rather than interfering with, the still weak capacities and values of the child. Acceptance of this presumption is compatible with the concession that it should be suspended or treated cautiously if the apparent wishes of the child diverge from the parental judgment.

Any values, preferences and life plans we can ascribe to the child are closely continuous and connected with, if not identical to, the parents' vision of their child's life. As described earlier, Whiting (3.1.1) and Brink (3.1.2) would say that there is interpersonal psychological continuity, or something close to it.⁵⁶² When making decisions for the child, such as decisions based on revelatory values, the parents are not deciding on behalf of someone else with whom they do not share a close and intimate connection. In this respect, the child shares features with the parents' own future selves: the parents shape the lives, preferences and views of these selves in a privileged way and in light of their current conception of their life stories. The same relations are plausibly understood as justifying parental decision-making authority, including revelatory assessments, on behalf of their child.

I conclude that there is a better application of the revelation approach than the one discussed by Shupe. It is reasonable for parents to evaluate the revelatory options on behalf of the child in light of their own values. I have left open which outcome this favors in a given case, but defended the revelation approach as one eligible tool for guiding decisions.

⁵⁶²See also D. Shoemaker 2007, p. 347.

7.7 Conclusion

Shupe argues that the revelation approach is unsuitable for making transformative choices for others. When deciding for ourselves, we are entitled to choose based on the value we assign to the revelation of novel experiences and perspectives. When deciding for others this seems far less appropriate. If anything, we should attempt to bring about beneficial consequences for the individual on whose behalf we are deciding. Focusing on revelation is inappropriate.

I have argued that Paul's revelation approach can counter these objections. The alleged *parent-focus* of the revelation approach need not be problematic (Strategy 1). And expected *consequences* can be factored into the revelatory values one assigns to the options (Strategy 3). Alternatively, one can deny that the revelation approach applies in a given instance of transformative choice for others, e.g., because the parents do not face a decision-problem that will be transformative for them (Strategy 2), or—if the evidence on the options' consequences is clear enough—because reflections about revelation are unnecessary to make a choice (Strategy 4). Moreover, it is true that applying the revelation approach in the course of *surrogate decision-making* is incredibly difficult. But this difficulty stems from more general challenges with surrogate decision-making, not from issues with the revelation approach in particular (Strategy 5). Uncertainty about the consequences, and the fact that the child has never been competent, are aspects that make the surrogate's decision-problem particularly hard.

Shupe raises her objections against a particular reading of the reve-

lation approach: that the parents shall focus on *their* revelatory values. I suggested and defended a different interpretation of what the approach requires (7.6): the parents shall evaluate the revelatory values *for their child*. When deliberating on behalf of the child, they are entitled to evaluate the options relative to their own values and views. Given their parental role and their intimate relation to their child, it is reasonable to presume that this does not interfere with a capacity for self-determination on the side of the child, insofar as it can be ascribed reflective capacities at all. Until then, parents have an entitlement, interest and right to raise their child in accordance with their values.

What remains is a question about consent not just in surrogate decision-making, but also in intrapersonal transformative choice. When being asked to choose amongst transformative treatments and treatments of transformative disease, the decision-maker is faced with options for which she cannot determine a subjective value, at least not through projective imagination the outcomes. If so, how can we make sense of seeking informed consent from individuals, and why do we not leave these intractable yet impactful decisions to the experts?

Chapter 8

Transformative Choice And Informed Consent

The characteristic feature of transformative choices is that at least one of the options is epistemically and/or personally transformative. We thus cannot assign subjective values and identify the best option.

One example is the cochlear implant case. It is impossible for the parents to imagine the transformative experience, i.e. receiving an implant that turns them from deaf to hearing. Hearing parents and deaf child have very different perspectives on the world. As such, the parents are radically ignorant of what the options of the decision problem will be like for their child. One worry is that the parents are in no position to make more than an uninformed guess about which revelation they would value most if they were in their child's situation.

The same problem arises with transformative choices for oneself. Paul thus points out that independently of the difficulties of surrogate decision-making, there is something odd about seeking informed consent. We cannot imagine what transformative experiences and their consequences are like. This means that a range of medical interventions and treatment regimes that are transformative or influence a disease with transformative consequences⁵⁶³ cannot be properly evaluated. We cannot know what experiencing them will be like, and are thus unable to make a rational decision. This raises the question why we are seeking informed consent for such interventions. For example, consider a treatment option that

⁵⁶³Carel et al. (2016, pp. 1152-3) argue that serious illness is transformative in Paul's terms. Illness yields a range of experiences for which we antecedently do not know what it is like to have them: receiving diagnoses, feeling symptoms, undergoing procedures, and coping with their effects. Moreover, the authors see the fact that patients with serious conditions most often undergo a process of adaptation such that they rank their well-being higher than healthy control subjects when asked to imagine living with that particular condition as evidence that they underwent personally transformative experiences.

might leave you disabled, and compare it with a different option that is less risky but also promises to be less effective. Paul writes:

“The basic idea is that you, as the patient, are supposed to evaluate the different outcomes, determine their values along with the level of risk you can tolerate as a result, and control the decision-making on that basis. But, crucially, then you need to know how you think you’ll respond to different possible experiential outcomes. And in this context, you are unable to get the information you need. [...] If you cannot determine your values and considered preferences for the outcomes, you cannot rationally determine the act with the highest expected value. The point [...] raises the question: what are we doing when we give our informed consent? [...] [W]hy think that an ‘informed’ consent is doing what it is supposed to do, that is, why think that the agent is really being given the opportunity to make a choice while being informed in the relevant way? [...] If the justification of informed consent is rooted in a person’s ability to understand her values and preferences concerning different possible outcomes, transformative choices pose a serious challenge.”⁵⁶⁴

It is not entirely clear whether Paul is merely making an observation or also suggesting changes. The observation is that problems of transformative choice can arise when informed consent is needed. If a transformation is looming, one cannot know how one will respond to the outcome. The informedness of the consent is thus inevitably limited. However, it sounds as if Paul is suggesting something more: that because of these issues, the importance and value of informed consent is deflated. Patients do not know what the outcomes of transformative experiences will be like, let alone have the medical expertise to evaluate treatment options. We should thus take their inputs less seriously.⁵⁶⁵

What *are* we doing when we seek informed consent? Does transfor-

⁵⁶⁴Paul 2014, p. 138.

⁵⁶⁵In line with this reading, Paul quotes Levy (2014), who argues that because of several biases and limitations in our ability to reason, *constraining* informed consent procedures would *promote* the good that informed consent is supposed to protect.

mation preclude being informed in the relevant way? Is the point of informed consent undermined by transformative experiences? I will now discuss three purposes of informed consent, and argue that transformative experience does not stand in the way of informed consent playing its role. We should thus disagree with Paul insofar as she is suggesting that transformative experience deflates the importance of informed consent.

8.1 Protection

One purpose of seeking informed consent is the protection of the patient or research subject.⁵⁶⁶ It is intended to prevent interference with an individual's liberty and freedom. J.S. Mill claims that individual liberties should be protected, that individuals know best what is good for them, and that interference with their pursuit of the good is acceptable only to prevent harm to others.⁵⁶⁷ Liberties provide the freedom of action to live in accordance with one's own personal views, taste and character.

Informed consent can further be seen as designed to prevent harm. Physicians' and researchers' interactions with patients are sometimes guided by interests of their own. A consent requirement is first stated in the *Nuremberg Code* which was formulated in view of experiments carried out on concentration camp inmates.⁵⁶⁸ Understood in this way, the function of consent is to prevent assault, coercion, and exploitation.

Moreover, consent procedures are valuable towards maintaining trust between patients and healthcare providers, which sustains numerous ben-

⁵⁶⁶Eyal 2012, sections 2.1, 2.3-4, 2.6.

⁵⁶⁷Mill 1859.

⁵⁶⁸See for example Manson and O'Neill 2007, ch. 1

official practices and institutions that would otherwise be in danger.⁵⁶⁹

I think it is relatively obvious that the protective functions of informed consent are not endangered by transformative experience. Part of what we are doing with informed consent procedures is protecting patients and subjects. They deserve information about available evidence, and are free to refuse treatment and participation. In order to be able to exercise one's liberty and decline harmful and exploitative measures, it is not necessary to know in detail how one would respond. Consent procedures can prevent interferences and harms even if the options presented in the consent process concern transformative options.

8.2 *Autonomy*

Informed consent is, ideally, a means to exercise autonomy.⁵⁷⁰ Taking autonomous choices seriously is necessary for respecting the person who defines and creates her life through them. For example, Harris writes:

“By shaping our lives for ourselves we assert our own values and our individuality. Our own choices, decisions, and preferences help to make us what we are, for each helps us to confirm and modify our own character and enables us to develop and to understand ourselves. So autonomy, as the ability and the freedom to make the choices that shape our lives, is quite crucial in giving to each life its own special and peculiar value. It is because we accept that the meaning, purpose, and indeed the distinctive uniqueness of an individual's life is given largely by acts of self definition and self creation that we are concerned to protect those attempts at self creation even where we are convinced that they are misguided or even self harming. [...] Informed consent is a dimension of respect for persons in that it is through consenting to things that affect us that we make those things consistent with our own values. When we consent to what others propose we make their ends and

⁵⁶⁹O'Neill 2002, p. 145.

⁵⁷⁰Eyal 2012, sect. 2.2.

objectives part of our own plans; so far from being merely the instruments of others we incorporate their plans and objectives into our own scheme of things and make them in that sense our own.”⁵⁷¹

Whether transformative experiences pose obstacles to informed consent understood as a means for exercising autonomy depends on what exactly one means by autonomy. There are many different accounts in the literature. I shall now discuss the following selection.⁵⁷²

8.2.1 Dworkin

G. Dworkin proposes that autonomous persons are both independent and authentic.⁵⁷³ Independence requires freedom from force, while authenticity in Dworkin’s sense is determined by the attitude one takes towards one’s first-order desires. Authentic persons act on desires which they adopt on the basis of higher-order preferences.⁵⁷⁴

Whether transformative choices can be authentic in this sense depends on what kind of second-order preferences are required. It seems I can form such preferences perfectly well by reflection that does not involve projective imagination of experiencing the considered option. Unless autonomy requires a specific kind of second-order preference that essentially depends on imaginative projection, transformative choices can be autonomous in this sense.

⁵⁷¹Harris 2003, p. 11.

⁵⁷²Authors differ on whether persons or actions are bearers of the property *autonomous*. For the sake of simplicity, my phrasings will assume that an action is autonomous iff it is carried out by an autonomous person.

⁵⁷³Dworkin 1976.

⁵⁷⁴Dworkin here refers to Frankfurt 1971. This is often perceived as a relatively demanding condition for autonomy. It seems we autonomously stop at red traffic lights without reflecting about our higher-order desires; see for example Faden and Beauchamp 1986, p. 264.

8.2.2 Faden And Beauchamp

Faden and Beauchamp understand informed consent as an autonomous authorization.⁵⁷⁵ They define:

“X acts autonomously only if X acts

1. intentionally,
2. with understanding, and
3. without controlling influences.”⁵⁷⁶

Rough glosses of the conditions include the following. An act is intentional if it “correspond[s] to the actor’s conception [...] of the act in question”⁵⁷⁷. Acting with understanding requires appropriate beliefs about “(1) the nature of the action, and (2) the foreseeable consequences and possible outcomes”⁵⁷⁸. Controlling influences are absent if nothing “deprive[s] the actor in any way of willing what he or she wishes to do or to believe.”⁵⁷⁹

As Faden and Beauchamp mention, acting intentionally seems to involve a form of cognitive engagement with the act and its consequences whose precise nature is difficult to spell out. I merely note that failure to imagine projectively and other features of transformative experience are unlikely to stand in the way of choosing intentionally. For one thing, imaginative projection would be a surprisingly strong condition for intentional action. Moreover, Paul never says that transformation precludes acting intentionally; becoming a parent might be transformative, but it is not the case that becoming a parent is always a non-intentional act.

⁵⁷⁵[Ibid.](#), pp. 277-80.

⁵⁷⁶[Ibid.](#), p. 238.

⁵⁷⁷[Ibid.](#), p. 242.

⁵⁷⁸[Ibid.](#), p. 251.

⁵⁷⁹[Ibid.](#), p. 258.

Similar remarks apply to the other conditions. Presumably, acting with understanding can be achieved by recognition of merely third-personal data, and need not involve imaginative projection into the outcomes. And whether controlling influences are present seems entirely independent from whether or not one faces transformation.

I conclude that given Faden's and Beuachamp's definition, transformative choices can be autonomous.

8.2.3 Savulescu

Savulescu maintains that only rational desires are expressive of autonomy. Subject P rationally desires A over B if the desire is based on

“(1) knowledge of relevant, available information concerning each of the states of affairs A and B, (2) no relevant, correctable errors of logic in evaluating that information, and (3) vivid imagination by P of what each state of affairs would be like for P.”⁵⁸⁰

Very much in Paul's spirit, Savulescu stresses the importance of vivid imagination for self-realization:

“[s]elf-determination is an active process of actually determining the path of ones life. In order to judge what is best for himself, P must think and imagine what it would be like for her if A and B obtained, and what the consequences, at least in the short term, of each of these would be for her. Thus, not only must P know what A and B are like, but she must also imagine what A and B would be like for her. I call this vivid imagination.”⁵⁸¹

He admits that P could be seen as *choosing* A even without vividly imagining alternative B. But without vivid imagination, the choice would fall short of self-realization.

⁵⁸⁰Savulescu 1994, p. 194.

⁵⁸¹*Ibid.*, p. 196.

Savulescu anticipates the objection that (3) is too strong. Captain Cook and Columbus could not imagine the outcomes of their explorations, yet embarked on them autonomously. In response, Savulescu clarifies that he only requires decision-makers to imagine outcomes

*“as far as possible. [...] One can autonomously choose to explore the unknown. However, one must gather as many facts as possible about the unknown, if one is to choose to explore it autonomously.”*⁵⁸²

If Paul is right about transformative experiences, thinking as far as possible means thinking not far at all. Our ways to vividly imagine transformative outcomes are severely limited. Savulescu is aware of a related challenge:

*“people may have a good idea about treatment decisions relating to familiar problems, like having a cold, a sprained or broken limb. But, as treatment decisions pertain to more unfamiliar states, people will have more difficulty in imagining what these states will be like. The evidence is that, prior to experiencing them, they will systematically underrepresent their utility. [...] In conclusion, the process of vivid imagination so necessary for evaluation is no easy or straightforward process to engage in. It involves not only the provision of much information, but the overcoming of several innate psychological hurdles.”*⁵⁸³

The hurdles Savulescu has in mind concern cognitive biases, such as loss aversion, discounting of the far future, and the fact that the quality of experiential states is determined by how they contrast with preceding states. The point here is slightly different from Paul's: people make mistakes when assigning subjective values to outcomes. In contrast, Paul's point was that people facing transformative experiences are unable to assign any subjective values whatsoever.

⁵⁸²[Ibid.](#), p. 197.

⁵⁸³[Ibid.](#), pp. 207-8.

To me, it seems slightly *ad hoc* on Savulescu's side to require imagining *as far as possible* for autonomy, but also to accept third-personal evidence as a surrogate if the outcome is unimaginable. However, I am not concerned with whether Savulescu's view is appropriate. What matters is that in his view, considering third-personal evidence can suffice for choosing autonomously and thereby exercising self-determination. Transformative choices can be autonomous in this sense if the decision-maker reflects on the outcomes by means of such data.

8.2.4 DeGrazia

DeGrazia defines autonomous action as follows:

“[subject] A autonomously performs intentional action X if and only if (1) A does X because she prefers to do X, (2) A has this preference because she (at least dispositionally) identifies with and prefers to have it, and (3) this identification has not resulted primarily from influences that A would, on careful reflection, consider alienating.”⁵⁸⁴

While the first two conditions overlap with Dworkin-style accounts, (3) is worth a separate discussion. With alienating influences, DeGrazia means things like coercion, force, or chance. Such influences might already stand in the way of satisfying (2), but supposing that (2) is satisfied somehow, they certainly conflict with (3): the decision-maker identifies in some sense with the preference as required by (2), but the way she does so is defective, or at least precludes autonomy.

Although beyond DeGrazia's intentions, we could see transformative experience as potentially conflicting with (3). For example, suppose prior

⁵⁸⁴DeGrazia 2005, p. 102. Italics removed.

to becoming a vampire, I identified with my preference to become one on the grounds of imagining how exciting and fulfilling vampire life would be. After reading Paul's work, I realize that I was ill-advised in identifying with the first-order preference on such grounds: I had no idea what vampire life would be like. In a sense, my identification was subject to an alienating influence, and my action not autonomous in DeGrazia's sense.

I am convinced that considerations along these lines do not establish that a transformative choice must be non-autonomous. At best, they demonstrate that if I have thoughts about what it's going to be like to transform, these can turn out to be mistaken or unfounded, and thus could be considered alienating. What this does not establish is that if I follow Paul's recommendation to *refrain* from attempts to base my decision on what I think the transformation will be like, I am subject to an alienating influence. In fact, Paul's very own revelation approach ties my choice closely to my core values and preferences without requiring imaginative projection into the outcomes. Even if a genuinely new kind of experience is looming, my choice to undergo it need not be subject to alienating influences.

Another of DeGrazia's points deserves attention. While discussing advance directives, he notes that their authority rests on some idealising assumptions about the individual who formulates them. She should be fully competent when she formulates the directive, and foresee the circumstances for which she forms and articulates her preferences.⁵⁸⁵ It can turn out that these conditions were not satisfied in a given case, which

⁵⁸⁵*Ibid.*, pp. 163-4 For further literature, see *ibid.*, fn. 9.

would negatively affect the authority of the directive in question. Specifically, an individual's ignorance or confusion about potential outcomes can affect the authority of her advance directives. For example, it is difficult to know or imagine what it will be like to be in a demented state and have a drastically different mental life.⁵⁸⁶ If a person had false ideas about what it would be like, and formulated its directive on the basis of these ideas, we should treat her directive with caution.

DeGrazia thinks that the force of this challenge is limited. The realization that an individual consented while being severely misinformed does undermine the moral authority or value of her consent, but this would merely be a local suspension of the general presumption that consent, even if expressed through directives, succeeds in articulating the preferences of the consenting individual.⁵⁸⁷ The risk I see for this response is that Paul sees the imaginative shortcomings she describes as *principled* and *deep* obstacles for the individual's consent being expressive of her preferences. Individuals' ideas about transformative treatments and diseases are *always* off the mark. At least for transformative experience, DeGrazia's 'general presumption' would never be justified.

My response parallels the remarks on Faden and Beauchamp's understanding condition. In order for consent to be expressive of one's preferences, it is not necessary that one succeeds in imagining the possible outcomes from the inside. If considering third-personal data is the best I can do for determining what I would prefer to happen to me in post-transformative states, then this is an appropriate basis for my de-

⁵⁸⁶DeGrazia 2005, p. 186.

⁵⁸⁷*Ibid.*, pp. 188-9.

liberation. The decision cannot be authentic in Paul's sense, given that I cannot now know what it's like to be severely demented, but it need not be authentic in order to be expressive of my autonomy.

This concludes my discussion of whether transformative choices complicate the exercise autonomy and hence one of the primary aims of informed consent. The surveyed accounts differ in their details. Characteristic features of transformative experiences could be seen as conflicting with authenticity in Dworkin's sense, intentionality, understanding, rational desires which require vivid imagination, or count as alienating influences. However, even for the more demanding accounts of autonomy, there is good reason to think that the transformative character of a choice does not in fact preclude its autonomy. When deciding for ourselves, autonomy can be exercised despite the imaginative limitations that constrain thinking about transformative options.

The insight that transformative choices and autonomy are compatible does not yet establish that informed consent plays its role. It remains a possibility that in order for informed consent to fulfil its purpose, a special kind of autonomous decision-making is needed for which the ability to imagine projectively is indispensable.

8.3 Self-Realization

Maybe informed consent in connection with transformative experience can be autonomous, but nevertheless fail to advance a related aim. Recall Harris' emphasis on how through autonomous choice, we assert our values, define our individuality, and steer our lives. Here, an obvious

connection can be drawn to self-constitution accounts of personal identity (3.2). Through consenting, we make the plans of others consistent with our own. The meaning, purpose and uniqueness of a life arises to an important extent from acts of self-definition and self-creation. It is plausible that one important point of informed consent is enabling such self-creative acts, and that the inability to projectively imagine transformative outcomes precludes that consent plays this role.

In order to respond to this worry, I repeat two insights: first, even the best transformative choices cannot rely on assignments of subjective values to the outcomes based on what it would be like for us to experience them from our first-person perspective. Secondly, this does not prevent transformative choices from playing a crucial role in shaping our lives:

“[y]our own choices involving transformative experiences, that is, your transformative choices, allow you to causally form what it will be like to be you in your future. In this sense, you own your future, because it is you who made the choice to bring this future—your very own future self—into being.”⁵⁸⁸

Paul points out that we need to make a concession. This process of self-creation is much less informed than we think. We do not know what it will be like to transform, and there is no way to assess in advance and first-personally how much pleasure, satisfaction, etc. a transformative outcome brings. This is why epistemic humility about our choices is required.

However, this does *not* mean that we cannot make well-motivated transformative choices. I understand the revelation approach as one way in which we can evaluate transformative options in light of our current

⁵⁸⁸Paul 2014, p. 18.

values and preferences, despite our inability to imagine what they will be like. It gives a sense in which we can make choices that change the shape and direction of our lives if not authentically (in Paul's sense, as being based on projective imagination), then still in light of who we are and what matters to us.

Enabling self-realization might be one of the purposes of informed consent procedures. Transformative choices are important means for steering our lives. The prospect of transformative treatments and diseases thus does not undermine the point of informed consent—especially if the decision-maker imagines the options first-personally as far as possible, considers third-personal evidence, weighs risks in accordance with her values, and most importantly decides based on how she now wishes to direct her life.

8.4 Conclusion

Paul worries that when illness and/or treatments are transformative, and transformative experience is understood properly, it is not clear what we are doing when we ask for informed consent. The decision-maker is radically ignorant of what is at stake and what the options are like, and thus seems to be bound to rely on a defective decision-making process which informed consent procedures declare authoritative.

I considered three purposes of informed consent procedures: protection against harms, enabling the individual to exercise autonomy, and facilitating self-realization. Although substantive accounts of the notions vary, I argued that the transformative character of a choice is not a prin-

ciplined obstacle to informed consent advancing these aims. Whether an individual makes a transformative choice for herself or on behalf of someone else; limitations in the ability to imagine what it is like to undergo the experience do not deflate the importance of her judgement.

This gives us a *prima facie* reason to leave the status and significance of informed consent unchanged. But I admit that this conclusion remains provisional. Further purposes of informed consent might become salient, and it is possible that the transformative character of some experiences and treatments stand in their way. Moreover, for the purposes I considered, alternative accounts, e.g., of autonomy, might become salient and raise similar problems. Until then, we should not take the transformative character of a patient's or research subject's decision as indication that informed consent is less likely to be "doing what it is supposed to do"⁵⁸⁹.

⁵⁸⁹Paul 2014, p. 138.

Conclusion

I described the complex versus simple distinction as it is drawn in the literature (chapter 1). I focused on what proponents of either side take to be advantages of their view, and what they think is wrong with its competitor.

According to objections primarily from simple theorists, the complex view must see person stages at different times as distinct, and reduces personal identity to relations of continuity and connectedness which by themselves are unimportant. Objections like these motivate the extreme claim: the complex view fails to give reasons for future-directed concern (chapter 2). I argued that the extreme claim is not a reason to discredit the complex view. The difficulty it highlights is not addressed more convincingly by the simple view. And the relations of continuity and connectedness to which the complex view refers are appropriate objects of concern. In particular, absence of a deep further fact does not diminish the value of ordinary survival as understood by the complex view.

I then characterized two positive accounts of how concern is justified (chapter 3). According to the first, relations of continuity and connectedness provide reasons for concern. According to the second, acts of self-constitution, either of ourselves as agents, or of our narrative identities, are what relates us to future selves such that their experiences and well-being matter to us now.

Once we realize that personal identity and relations of practical importance need not go together, several problems with transplant intuitions emerge (chapter 4). These intuitions are supposed to help us to choose between psychological and bodily criteria of personal identity, the leading instances of the complex view. I did not deny that transplant

intuitions reflect what we care about when psychological and bodily continuities come apart. But I argued that regarding personal identity, they provide only circular evidence, seem uninformative if quasi-attitudes are possible, and diverge inexplicably across cases. Transplant intuitions thus do not bring animalism or any other theory under pressures.

It then became apparent that I had made a concession: I took for granted that there is a substantive complex versus simple distinction. Until here, I argued that even if we buy into the distinction, it does not have the practical significance which many attribute to it. But in (chapter 5), I argued that the complex versus simple distinction is actually confused. Paradigm cases of the simple view can satisfy the conditions thought to be distinctive of the complex view. And Noonan makes complexity and simplicity turn on features whose relevance for the distinction is questionable. As a result, the distinction proved to be unhelpful for the *classification* and *endorsement* of positions.

After having provided this diagnosis, I owed suggestions on how, if not in terms of the complex versus simple distinction, we should we frame theories and debates about personal identity. I characterized three examples for notions that could replace the distinction (chapter 6): ontological dependence, fundamentality, and essentialism. Replacing the complex versus simple distinction allows us to understand and compare positions better, to defend them against charges of elusiveness, and to prevent us from making misguided commitments.

One insight from the foregoing discussions was that what matters for a range of practical activities, most notably future-directed concern, is less obviously connected to views of numerical identity over time than

many think. Turning to Paul, a complementary insight became apparent: even if we keep *numerical* identity over time fixed, a person undergoes *qualitative* changes throughout her lives. The question arises how she can ever rationally commit to undergo transformative experiences, i.e. experiences that fundamentally change her epistemic perspective and/or core preferences (chapter 7). Paul proposes that becoming someone else in this sense is rational if we antecedently value the discovery of novelty more than the status quo. I put this proposal to the test against objections by Shupe. She argues that *numerical* identity and distinctness *is* fundamentally important when justifying transformative choices: we can grant Paul that her approach is satisfactory when making transformative choices for ourselves. Still, the account breaks down when the transforming individual is numerically distinct from the decision-maker who evaluates the options. The illustrative case is one in which parents need to decide whether their deaf child shall receive a cochlear implant. I provided a reading of Paul's approach that differs from Shupe's, and argued that it gives a plausible account of how the parents can make a rational choice in this case. The approach cannot—and should not be expected to—make the parents' incredibly hard decision-problem easier, given the nature of the case and the uncertainties that surround it. What matters is that Paul's approach is not bound to make demands that are too parent-focused, ignore consequences, and/or complicate surrogate decision-making.

Paul's insights on transformative experience lead to a problem. The decision-maker who faces transformative options is radically ignorant of how she will respond to the outcomes. This applies to clinical and re-

search contexts, where patients and subjects often need to weigh transformative options and provide or refuse informed consent. If it is inevitable that their judgement is poorly informed, what are we doing when we seek informed consent? I argued that the failure to projectively imagine transformative outcomes does not stand in the way of three important purposes of informed consent: protection from interference and harm, enabling the exercise of autonomy, and allowing self-realization (chapter 8). Without further argument, the transformative character of treatments and disease does not justify diminishing the importance of informed consent.

In light of these insights, we can conclude that the interdependence between personal identity and practical reason is overstated. I did not set out to survey each and every possible connection between the two topics. My goal was to evaluate paradigmatic instances where philosophers and common sense suggest that personal identity constrains how we should reason and care, or vice versa. My conclusions suggest that in these instances, the two domains are in fact neutral to each other. Practical concerns do not give us a reason to favour the simple view. The complex view does not fly in the face of practical concerns. What matters practically might be related and contribute to personal identity, but it is something else. Intuitions that supposedly favour particular positions about personal identity are really about practical concerns, which might often be related to personal identity but can come apart from it. In fact, numerical identity does not preclude a justificatory challenge about certain intrapersonal choices: if a transformation is looming, theories of personal identity and standard decision theory struggle to explain how the

decision-maker can choose rationally. The fact that the decision-maker is ignorant of how she will change does not diminish the importance of informed consent procedures that allow her to make autonomous choices expressive of her personhood about how to steer her life.

I close with two further points. First, my discussion highlights an aspect which complicates debates about personal identity and its relation to practical concerns. Authors differ with regards to a fundamental assumption about personhood and personal identity. On the one hand, there is a broadly Lockean tradition. It ties personal identity to forensic and other practically relevant kinds of unity, e.g., being one and the same unit for the aggregation of sacrifices and benefits, or being one and the same subject in a sense that justifies concern. Where a person begins and ends is somehow connected to practical matters. As Eklund puts it, these positions endorse the *moral relevance* assumption about persons.

On the other hand, there is the view that all these matters are entirely irrelevant to numerical personal identity over time. In Olson's terms, these practical matters pertain to *being the same person*, but it is an open question whether personal identity depends on them (Olson denies it). An analogous thought is present in Schechtman's distinction between personal identity in the *reidentification* sense, and the narrative identity of persons. Practical concerns speak to the latter and vice versa. But practical concerns are orthogonal to *reidentification*.

What this divide suggests is that part of why it is so difficult to say something principled about the interdependence between personal identity and practical reason is that much hinges on how practice-laden one understands the notion of personhood to be. We are bound to talk

past each other and engage in verbal disagreements if differences in the conceptual and terminological departure points at such a fundamental level remain unrecognized. We are also bound to proceed on the basis of bad evidence, as the discussion on transplant intuitions has shown. My demand is thus that we are transparent about which side of the divide just characterized we endorse, and engage in substantive debates about this topic.

Finally, I close by explaining how my claims relate to Parfit's work on personal identity. On the one hand, the insights summarized above bear some obvious parallels to his position. The diagnosis of a lack of straightforward interdependence between personal identity and practical reason resonates with the idea that identity is not what matters. We can have all we care about without remaining numerically identical. Indeed, I argued that the grounds for concern need not mention identity.

On the other hand, there are several points on which I disagree with Parfit. The most obvious is the complex versus simple distinction which he takes to be the foundation of his work. I have pointed at several problems with the distinction. I am convinced that the distinction should be seen as largely irrelevant to his claims about personal identity and what matters, which require independent motivations. Parfit is surely entitled to draw on puzzle cases like fission, the analysability of personal identity, the divisibility of persons, etc. But if I am right, all these resources are available to simple theorists, too.

Moreover, I denied that caring about relations of continuity and connectedness means we should care less about death. Identity and deep further facts may not necessarily matter. My point was that absence of

a deep further fact does not settle what matters in which way. What is left may still matter a great deal. I accept that it is permissible to feel liberated and care less in the way Parfit suggests. But I maintained that Parfit's descriptive picture of personal identity does not provide a motivation, let alone a rational requirement, to care differently than one did back in the glass tunnel.

Bibliography

- Adams, R. M. (1989). “Should Ethics be More Impersonal? A Critical Notice of Derek Parfit, *Reasons and Persons*”. In: *The Philosophical Review* 98.4, pp. 439–484.
- Baker, L. R. (2008). “Big-Tent Metaphysics”. In: *Abstracta* (Special Issue on Eric Olson’s *The Human Animal*), pp. 8–15.
- Barnbaum, D. (1999). “Interpreting Surrogate Consent using Counterfactuals”. In: *Journal of Applied Philosophy* 16.2, pp. 167–172.
- Barnes, E. (2012). “Emergence and Fundamentality”. In: *Mind* 121.484, pp. 873–901.
- (2015). “What You Can Expect When You Don’t Want to be Expecting”. In: *Philosophy and Phenomenological Research* 91.3, pp. 775–786.
- Barnett, D. (2008). “The Simplicity Intuition and Its Hidden Influence on Philosophy of Mind”. In: *Noûs* 42.2, pp. 308–335.
- (2010). “You are Simple”. In: *The Waning of Materialism*. Ed. by R. C. Koons and G. Bealer. Oxford: Oxford University Press, pp. 161–173.
- Beauchamp, T. L. and J. F. Childress (2013). *Principles of Biomedical Ethics*. 7th edition. Oxford: Oxford University Press.
- Beck, S. (1998). “Back to the self and the future”. In: *South African journal of philosophy* 17.3, pp. 211–225.
- (2014). “Transplant thought-experiments: two costly mistakes in discounting them”. In: *South African Journal of Philosophy* 33.2, pp. 189–199.
- (2015). “The extreme claim, psychological continuity and the person life view”. In: *South African Journal of Philosophy* 34.3, pp. 314–322.
- Belzer, M. (1996). “Notes on Relation R”. In: *Analysis* 56.1, pp. 56–62.

- Blackburn, S. (1997). "Has Kant Refuted Parfit?" In: *Reading Parfit*. Ed. by J. Dancy. Oxford: Blackwell.
- Blanke, O. (2012). "Multisensory brain mechanisms of bodily self-consciousness". In: *Nature Reviews Neuroscience* 13.8, pp. 556–571.
- Blustein, J. (1999). "Choosing for Others as Continuing a Life Story: The Problem of Personal Identity Revisited". In: *The Journal of Law, Medicine & Ethics* 27.1, pp. 20–31.
- Bottani, A. (2013). "Review of: Georg Gasser and Matthias Stefan (eds), Personal Identity: Complex or Simple?" In: *dialectica* 67.4, pp. 611–617.
- Braddon-Mitchell, D. and K. Miller (2004). "How to Be a Conventional Person". In: *The Monist* 87.4, pp. 457–474.
- Brink, D. O. (1997a). "Rational Egoism and the Separateness of Persons". In: *Reading Parfit*. Ed. by J. Dancy. Oxford: Blackwell, pp. 96–134.
- (1997b). "Self-Love and Altruism". In: *Social Philosophy and Policy* 14, pp. 122–157.
- (2017). "Special Concern and Personal Identity". Unpublished manuscript. URL: <http://davidobrink.com/work-progress/special-concern-prudence-and-personal-identity>.
- Brock, D. W. (1993). "Quality of Life Measures in Health Care and Medical Ethics". In: *The Quality of Life*. Ed. by M. Nussbaum and A. Sen. Oxford: Clarendon Press, pp. 95–132.
- (1999). "A Critique of Three Objections to Physician-Assisted Suicide". In: *Ethics* 109.3, pp. 519–547.
- Brueckner, A. (2009). "Endurantism and the Psychological Approach to Personal Identity". In: *Theoria* 75.1, pp. 28–33.
- Buchanan, A. E. and D. W. Brock (1990). *Deciding for Others: The Ethics of Surrogate Decision-Making*. Cambridge: Cambridge University Press.
- Butler, J. (1736). "Of Personal Identity". In: *Personal Identity*. Ed. by J. Perry. Berkeley: University of California Press 1975, pp. 99–105.
- Campbell, J. (2015). "L. A. Paul's Transformative Experience". In: *Philosophy and Phenomenological Research* 91.3, pp. 787–793.

- Carel, H., I. J. Kidd, and R. Pettigrew (2016). "Illness as transformative experience". In: *The Lancet* 388.10050, p. 1152.
- Chisholm, R. M. (1969). "The Loose and Popular and Strict and Philosophical Sense of Identity". In: *Perception and Personal Identity*. Ed. by N. Care and R. Grimm. Ohio University Press.
- (1976). *Person and Object: A Metaphysical Study*. Routledge 2004.
- (1986). "Self-Identity". In: *Roderick M. Chisholm*. Ed. by R. J. Bogdan. Dordrecht: Springer Science+Business Media, pp. 3–77.
- Correia, F. (2008). "Ontological dependence". In: *Philosophy Compass* 3.5, pp. 1013–1032.
- Correia, F. and B. Schnieder (2012). "Grounding: an opinionated introduction". In: *Metaphysical Grounding. Understanding the Structure of Reality*. Ed. by F. Correia and B. Schnieder. Cambridge: Cambridge University Press, pp. 1–36.
- Cowling, S. (2015). "Non-qualitative Properties". In: *Erkenntnis* 80.2, pp. 275–301.
- (2016). "Haecceitism". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2016. Metaphysics Research Lab, Stanford University.
- Crisp, T. and D. Smith (2005). "Wholly Present Defined". In: *Philosophy and Phenomenological Research* 71.2, pp. 318–344.
- Crone, K. (2017). "Strukturen der Identität und des Selbstverständnisses von Personen". In: *Deutsche Zeitschrift für Philosophie* 65.1, pp. 1–15.
- Daly, C. (2012). "Scepticism about grounding". In: *Metaphysical Grounding. Understanding the Structure of Reality*. Ed. by F. Correia and B. Schnieder. Cambridge: Cambridge University Press, pp. 80–100.
- DeGrazia, D. (2005). *Human Identity and Bioethics*. Cambridge: Cambridge University Press.
- Descartes, R. (1647). "Principles of Philosophy". In: *The Philosophical Writings of Descartes*. Ed. by J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Cambridge University Press 1985.
- Diekemper, J. (2009). "Thisness and Events". In: *The Journal of Philosophy* 106.5, pp. 255–276.

- Dummett, M. (1973). *Frege. Philosophy of Language*. New York: Harper & Row.
- Duncan, M. (2014). “A Challenge to Anti-Criterialism”. In: *Erkenntnis* 79.2, pp. 283–296.
- Dworkin, G. (1976). “Autonomy and behavior control”. In: *Hastings Center Report* 6.1, pp. 23–28.
- Ehrsson, H. H. et al. (2007). “Threatening a rubber hand that you feel is yours elicits a cortical anxiety response”. In: *Proceedings of the National Academy of Sciences* 104.23, pp. 9828–9833.
- Eklund, M. (2002). “Personal Identity and Conceptual Incoherence”. In: *Nous* 36.3, pp. 465–485.
- (2004). “Personal Identity, Concerns, and Indeterminacy”. In: *The Monist* 87.4, pp. 489–511.
- Eyal, N. (2012). “Informed Consent”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2012. Metaphysics Research Lab, Stanford University.
- Faden, R. and T. L. Beauchamp (1986). *A History and Theory of Informed Consent*. Oxford: Oxford University Press.
- Feldman, F. (1991). “Some Puzzles About the Evil of Death”. In: *The Philosophical Review*, pp. 205–227.
- Fine, K. (1994). “Essence and Modality”. In: *Philosophical Perspectives* 8, pp. 1–16.
- (1995). “Ontological Dependence”. In: *Proceedings of the aristotelian society*, pp. 269–290.
- (2001). “The Question of Realism”. In: *Philosopher’s Imprint* 1, pp. 1–30.
- Frankfurt, H. G. (1971). “Freedom of the Will and the Concept of a Person”. In: *The Journal of Philosophy* 68.1, pp. 5–20.
- Garrett, B. (1998). *Personal Identity and Self-Consciousness*. London and New York: Routledge.
- (2003). “Some Thoughts on Animalism”. In: *On Human Persons*. Ed. by K. Petrus. Frankfurt and London: ontos Verlag, pp. 41–45.
- (2006). *What is this Thing Called Metaphysics?* London and New York: Routledge.

- Gasser, G. (2013). "Personale Identität und die Sorge um sich selbst". In: *Personale Identität, Narrativität und Praktische Rationalität*. Ed. by G. Gasser and M. Schmidhuber. Münster: mentis.
- Gasser, G. and M. Stefan (2012). "Introduction". In: *Personal Identity: Complex or Simple?* Ed. by G. Gasser and M. Stefan. Cambridge University Press.
- Greene, J. D. (2008). "The Secret Joke of Kant's Soul". In: *Moral psychology* 3, pp. 35–79.
- Harris, J. (2003). "Consent and end of life decisions". In: *Journal of Medical Ethics* 29.1, pp. 10–15.
- Hawley, K. (2001). *How Things Persist*. Oxford: Clarendon Press 2004.
- (2004). "Borderline Simple or Extremely Simple". In: *The Monist*, pp. 385–404.
- Herdova, M. (2016). "Nothing to Fear: Swap Cases and Personal Identity". In: *Analytic Philosophy* 57.4, pp. 315–337.
- Higginbotham, J. (1995). "Tensed Thoughts". In: *Mind & language* 10.3, pp. 226–249.
- Hofweber, T. (2009). "Ambitious, Yet Modest, Metaphysics". In: *Metametaphysics. New Essays on the Foundations of Ontology*. Ed. by M. D. Chalmers D. and R. Wassermann. Oxford: Oxford University Press.
- Hofweber, T. and J. Velleman (2011). "How to endure". In: *The Philosophical Quarterly* 61.242, pp. 37–57.
- Hummel, P. (2012). "Personal Identity and Concern for the Future. Three Problems for Mark Johnston's Minimalism". M.Litt Dissertation. St Andrews and Stirling Graduate Programme for Philosophy.
- (2017). "Against the Complex Versus Simple Distinction". In: *Erkenntnis* 82.2, pp. 363–378.
- Johansson, J. (2007). "Non-Reductionism and Special Concern". In: *Australasian Journal of Philosophy* 85.4, pp. 641–57.
- Johnston, M. (1987). "Human Beings". In: *The Journal of Philosophy* 84.2, pp. 59–83.
- (1989). "Relativism and the Self". In: *Relativism, Interpretation and Confrontation*. Ed. by M. Krausz. Notre Dame: University of Notre Dame Press.

- Johnston, M. (1992). "Reasons and Reductionism". In: *The Philosophical Review* 101.3, pp. 589–618.
- (1993). "Objectivity Refigured: Pragmatism without Verificationism". In: *Reality, Representation, and Projection*. Ed. by J. Haldane and C. Wright. Oxford: Oxford University Press, pp. 85–130.
- (1997). "Human Concerns without Superlative Selves". In: *Reading Parfit*. Ed. by J. Dancy. Blackwell, pp. 149–179.
- (2010). *Surviving Death*. Princeton University Press.
- Joyce, R. (2000). "Darwinian Ethics and Error". In: *Biology and philosophy* 15.5, pp. 713–732.
- (2016). "Evolution, Truth-Tracking, and Moral Skepticism". In: *Essays in Moral Skepticism*. Oxford: Oxford University Press, pp. 142–158.
- Kahane, G. (2011). "Evolutionary Debunking Arguments". In: *Noûs* 45.1, pp. 103–125.
- Kind, A. (2004). "The Metaphysics of Personal Identity and our Special Concern for the Future". In: *Metaphilosophy* 35.4, pp. 536–553.
- Korsgaard, C. M. (1989). "Personal Identity and the Unity of Agency: A Kantian Response to Parfit". In: *Philosophy & Public Affairs* 18.2, pp. 101–132.
- Koslicki, K. (2012). "Varieties of ontological dependence". In: *Metaphysical Grounding. Understanding the Structure of Reality*. Cambridge: Cambridge University Press, pp. 186–213.
- Kovacs, D. M. (2016). "Self-made People". In: *Mind* 125.500, pp. 1071–1099.
- Kripke, S. (1972). *Naming and Necessity*. Cambridge, Massachusetts: Harvard University Press 1980.
- Langford, S. (2014). "Is Personal Identity Analysable?" In: *Acta Analytica* 29.3, pp. 309–316.
- Leibniz, G. W. (1686). "Discourse on Metaphysics". In: *Discourse on Metaphysics and Other Writings*. Ed. by P. Loftson. London: Broadview Editions.
- Levy, N. (2014). "Forced to be free? Increasing patient autonomy by constraining it". In: *Journal of Medical Ethics* 40, pp. 293–300.

- Lewis, D. K. (1976). "Survival and Identity". In: *The Identities of Persons*. Ed. by A. O. Rorty. Berkeley: University of California Press.
- (1986). *On the Plurality of Worlds*. Oxford: Basil Blackwell.
- Locke, J. (1690). *An Essay Concerning Human Understanding*. Ed. by P. Nidditch. Oxford: Oxford University Press 1975.
- Lowe, E. J. (1996). *Subjects of Experience*. Cambridge University Press.
- (2012). "The probable simplicity of personal identity". In: *Personal Identity: Complex or Simple?* Cambridge: Cambridge University Press, pp. 137–155.
- Mackenzie, C. (2008). "Introduction: Practical Identity and Narrative Agency". In: *Practical Identity and Narrative Agency*. Ed. by C. Mackenzie and K. Atkins. New York and London: Routledge.
- Madden, R. (2015). "The Naive Topology of the Conscious Subject". In: *Nôus* 49.1, pp. 55–70.
- (2016). "Human Persistence". In: *Philosopher's Imprint* 16.17.
- Madell, G. (1981). *The Identity of the Self*. Edinburgh: Edinburgh University Press.
- (2015). *The Essence of the Self. In Defense of the Simple View of Personal Identity*. Abingdon and New York: Routledge.
- Manson, N. C. and O. O'Neill (2007). *Rethinking Informed Consent in Bioethics*. Cambridge: Cambridge University Press.
- Martin, R. (1998). *Self-Concern. An Experiential Approach To What Matters in Survival*. Cambridge: Cambridge University Press.
- Merricks, T. (1994). "Endurance and Indiscernibility". In: *The Journal of Philosophy* 91.4, pp. 165–184.
- (1998). "There Are No Criteria of Identity Over Time". In: *Nôus* 32.1, pp. 106–124.
- (1999a). "Endurance, Psychological Continuity, and the Importance of Personal Identity". In: *Philosophy and Phenomenological Research* 59.4, pp. 983–997.
- (1999b). "Persistence, Parts, and Presentism". In: *Nôus* 33.3, pp. 421–438.
- (2000). "Perdurance and Psychological Continuity". In: *Philosophy and Phenomenological Research* 61.1, pp. 195–198.
- (2001a). *Objects and Persons*. Oxford: Clarendon Press.

- Merricks, T. (2001b). “Realism about Personal Identity over Time”. In: *Philosophical Perspectives* 15.
- Mill, J. S. (1859). “On Liberty”. In: *The Collected Works of John Stuart Mill*. Ed. by J. M. Robson. Toronto: University of Toronto Press.
- Miller, K. (2013). “Personal identity minus the persons”. In: *Philosophical Studies* 166.1, pp. 91–109.
- Nagasawa, Y. (2008). “Proxy Consent and Counterfactuals”. In: *Bioethics* 22.1, pp. 16–24.
- Nagel, T. (1979). *The Possibility of Altruism*. Princeton, New Jersey: Princeton University Press.
- Nichols, S. and M. Bruno (2010). “Intuitions about personal identity: An empirical study”. In: *Philosophical Psychology* 23.3, pp. 293–312.
- Ninan, D. (2009). “Persistence and the First-Person Perspective”. In: *Philosophical Review* 118.4, pp. 425–464.
- Noonan, H. W. (1989). *Personal Identity*. 2nd edition. Routledge 2003.
- (2007). “Identity eliminated”. In: *Analysis* 67.294, pp. 122–127.
- (2011). “The complex and simple views of personal identity”. In: *Analysis* 71.1, pp. 72–77.
- Nozick, R. (1981). *Philosophical Explanations*. Oxford: Clarendon Press.
- Olson, E. T. (1997). *The Human Animal: Personal Identity without Psychology*. Oxford University Press.
- (2007). *What Are We?* Oxford: Oxford University Press.
- (2012). “In search for the simple view”. In: *Personal Identity. Complex or Simple?* Ed. by G. Gasser and M. Stefan. Cambridge University Press, pp. 44–62.
- (2013). “The Person and the Corpse”. In: *The Oxford Handbook of Philosophy of Death*. Oxford: Oxford University Press, pp. 80–96.
- (2015a). “Animalism and the Remnant-Person Problem”. In: *Philosophical Perspectives on the Self*. Ed. by J. Fonseca and J. Gonalves. Bern: Peter Lang.
- (2015b). “On Parfit’s View That We Are Not Human Beings”. In: *Royal Institute of Philosophy Supplement* 76, pp. 39–56.

- (2017). “Personal Identity”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Summer 2017. Metaphysics Research Lab, Stanford University.
- (forthcoming). “For Animalism”. In: *The Blackwell Companion to Substance Dualism*. Ed. by M. A. Loose J. and J. P. Moreland. Oxford: Wiley.
- O’Neill, O. (2002). *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press.
- Oyowe, O. A. (2013). “Physical Continuity, Self and the Future”. In: *Philosophia* 41.1, pp. 257–269.
- Parfit, D. (1971). “Personal Identity”. In: *The Philosophical Review*, pp. 3–27.
- (1982). “Personal identity and rationality”. In: *Synthese* 53.2, pp. 227–241.
- (1984). *Reasons and Persons*. Oxford: Oxford University Press 1987.
- (1986). “Comments”. In: *American Journal of Bioethics* 96.4, pp. 832–872.
- (1995). “The Unimportance of Identity”. In: *Identity: Essays based on Herbert Spencer Lectures given in the University of Oxford*. Ed. by H. Harris. New York: Clarendon Press, pp. 13–45.
- (2007). *Is Personal Identity What Matters?* URL: <http://www.ammonius.org/assets/pdfs/ammoniusfinal.pdf>.
- (2012). “We are not Human Beings”. In: *Philosophy* 87.01, pp. 5–28.
- Paul, L. A. (2014). *Transformative Experience*. Oxford: Oxford University Press.
- (2015). “Précis of Transformative Experience”. In: *Philosophy and Phenomenological Research* 91.3, pp. 760–765.
- Paul, L. and K. Healy (forthcoming). “Transformative Treatments”. In: *Noûs*.
- Perry, J. (1976). “The Importance of Being Identical”. In: *The Identities of Persons*. Ed. by A. O. Rorty. Berkeley: University of California Press.
- Persson, I. (2005). *The Retreat of Reason: A Dilemma in the Philosophy of Life*. Oxford University Press on Demand.

- Pettigrew, R. (2016). "Transformative Experience, by L. A. Paul". In: *Mind* 125.499, pp. 927–935.
- Pollock, J. and J. Cruz (1986). *Contemporary Theories of Knowledge*. Lanham, Maryland: Rowman & Littlefield.
- Quine, W. V. (1976). "Worlds Away". In: *The Journal of Philosophy* 73.22, pp. 859–863.
- Rachels, S. and T. Alter (2005). "Nothing Matters in Survival". In: *The Journal of Ethics* 9.3-4, pp. 311–330.
- Raven, M. J. (2013). "Is ground a strict partial order?" In: *American Philosophical Quarterly* 50.2, pp. 193–201.
- Rayo, A. (2009). "Toward a Trivialist Account of Mathematics". In: *New Waves in Philosophy of Mathematics*. Ed. by O. Bueno and Ø. Linnebo. London: Palgrave Macmillan, pp. 239–260.
- Rea, M. C. and D. Silver (2000). "Personal Identity and Psychological Continuity". In: *Philosophical and Phenomenological Research*, pp. 185–193.
- Recanati, F. (2007). *Perspectival Thought: A Plea for (Moderate) Relativism*. Clarendon Press.
- Reid, T. (1785a). *Essays on the Intellectual Powers of Man*. Ed. by D. R. Brookes. University Park, Pennsylvania: Pennsylvania University Press 2002.
- (1785b). "Of Mr. Locke's Account of our Personal Identity". In: *Personal Identity*. Ed. by J. Perry. Berkeley: University of California Press 1975, pp. 113–118.
- Roca-Royes, S. (2011). "Essential Properties and Individual Essences". In: *Philosophy Compass* 6.1, pp. 65–77.
- Rosen, G. (2010). "Metaphysical Dependence: Grounding and Reduction". In: *Modality. Metaphysics, Logic and Epistemology*. Ed. by B. Hale and A. Hoffmann. Oxford: Oxford University Press, pp. 109–136.
- Savulescu, J. (1994). "Rational Desires and the Limitation of Life-Sustaining Treatment". In: *Bioethics* 8.3, pp. 191–222.

- (2009). *Refusing Cochlear Implants: Is It Child Neglect?* URL: <http://blog.practicaethics.ox.ac.uk/2009/07/refusing-cochlear-implants-is-it-child-neglect/>.
- Schaffer, J. (2010). “Monism: The Priority of the Whole”. In: *Philosophical Review* 119.1, pp. 31–76.
- Schechtman, M. (1996). *The Constitution of Selves*. Cornell University Press.
- (2014). *Staying Alive. Personal Identity, Practical Concerns, and the Unity of Life*. Oxford: Oxford University Press.
- Schnieder, B. (forthcoming). “Grounding and dependence”. In: *Synthese*.
- Seckler, A. B. et al. (1991). “Substituted Judgment: How Accurate Are Proxy Predictions?” In: *Annals of Internal Medicine* 115.2, pp. 92–98.
- Shoemaker, D. W. (2006). “Parfit’s ‘Argument from Below’ vs. Johnston’s ‘Argument from Above’”. In: *PEA Soup*. URL: http://peasoup.typepad.com/peasoup/2006/04/parfits_argumen.html.
- (2007). “Personal Identity and Practical Concerns”. In: *Mind* 116.462, pp. 317–357.
- (2016). “The Stony Metaphysical Heart of Animalism”. In: *Animalism. New Essays on Persons, Animals, and Identity*. Ed. by S. Blatti and P. F. Snowdon. Oxford: Oxford University Press, pp. 303–327.
- Shoemaker, S. (1963). *Self-Knowledge and Self-Identity*. Ithaca, New York: Cornell University Press.
- (1970). “Persons and their Pasts”. In: *American Philosophical Quarterly* 7.4, pp. 269–285.
- (1984). “Personal Identity: a Materialist Account”. In: *Personal Identity*. Ed. by S. Shoemaker and R. Swinburne. Oxford: Blackwell.
- (1985). “Critical Notice of Reasons and Persons”. In: *Mind* 94, pp. 443–53.
- (2012). “Against simplicity”. In: *Personal Identity. Complex or Simple?* Ed. by G. Gasser and M. Stefan. Cambridge University Press, pp. 123–136.
- Shupe, E. (2016). “Transformative experience and the limits of revelation”. In: *Philosophical Studies* 173.11, pp. 3119–3132.

- Sider, T. (1996). "All the world's a stage". In: *Australasian Journal of Philosophy* 74.3, pp. 433–453.
- (2001a). "Criteria of Personal Identity and the Limits of Conceptual Analysis". In: *Noûs* 35.s15, pp. 189–209.
- (2001b). *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford University Press.
- (2013). "The Evil of Death: What Can Metaphysics Contribute?" In: *The Oxford Handbook of Philosophy of Death*. Ed. by F. F. Bradley B. and J. Johansson. Oxford: Oxford University Press.
- Sidgwick, H. (1907). *The Methods of Ethics*. London: Macmillan.
- Snowdon, P. (2014). *Persons, Animals, Ourselves*. Oxford: Oxford University Press.
- Sosa, E. (1990). "Surviving Matters". In: *Noûs*, pp. 297–322.
- Strawson, P. F. (1962). "Freedom and Resentment". In: *Freedom and Resentment and Other Essays*. London and New York: Routledge 2008.
- Street, S. (2006). "A Darwinian Dilemma for Realist Theories of Value". In: *Philosophical Studies* 127.1.
- (2009). "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters". In: *Philosophical Issues* 19.1, pp. 273–298.
- Swinburne, R. (1974). "Personal Identity". In: *Proceedings of the Aristotelian Society*. Vol. 74, pp. 231–247.
- (1984). "Personal Identity: a Dualist Theory". In: *Personal Identity*. Ed. by S. Shoemaker and R. Swinburne. Oxford: Blackwell.
- Tahko, T. E. and E. J. Lowe (2016). "Ontological Dependence". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Winter 2016. Metaphysics Research Lab, Stanford University.
- Talbott, W. (2016). "Transformative Experience". In: *Analysis*, anw032.
- Unger, P. (1990). *Identity, Consciousness, and Value*. New York: Oxford University Press.
- Van Inwagen, P. (1990). *Material Beings*. Ithaca: Cornell University Press.
- (1997). "Materialism and the Psychological-Continuity Account of Personal Identity". In: *Noûs* 31.s11, pp. 305–319.

- Velleman, J. D. (1996). "Self to Self". In: *The Philosophical Review* 105.1, pp. 39–76.
- Wasserman, R. (2012). "Personal identity, indeterminacy, and obligation". In: *Personal Identity. Complex or Simple?* Ed. by G. Gasser and M. Stefan. Cambridge University Press, pp. 63–81.
- Whiting, J. (1986). "Friends and Future Selves". In: *The Philosophical Review* 95.4, pp. 547–580.
- (1999). "Back to 'The Self and the Future'". In: *Philosophical Topics* 26.1/2, pp. 441–477.
- Wielenberg, E. J. (2016). "Ethics and Evolutionary Theory". In: *Analysis* 76.4, pp. 502–515.
- Wierenga, E. (1983). "Proxy Consent and Counterfactual Wishes". In: *Journal of Medicine and Philosophy* 8.4, pp. 405–416.
- Williams, B. (1956). "Personal Identity and Individuation". In: *Proceedings of the Aristotelian Society*. Vol. 57. JSTOR, pp. 229–252.
- (1970). "The Self and the Future". In: *The Philosophical Review* 79.2, pp. 161–180.
- Williams, J. R. G. (2007). "The possibility of onion worlds: Rebutting an argument for structural universals". In: *Australasian Journal of Philosophy* 85.2, pp. 193–203.
- Wilson, J. M. (2014). "No Work for a Theory of Grounding". In: *Inquiry* 57.5-6, pp. 535–579.
- Wolf, S. (1986). "Self-Interest and Interest in Selves". In: *Ethics* 96.4, pp. 704–720.
- Wrigley, A. (2011). "The Problem of Counterfactuals in Substituted Judgement Decision-Making". In: *Journal of Applied Philosophy* 28.2, pp. 169–187.
- Zahavi, D. (2007). "Self and Other: The Limits of Narrative Understanding". In: *Royal Institute of Philosophy Supplements* 60, pp. 179–202.
- Zimmerman, D. (1998). "Criteria of Identity and the Identity Mystics". In: *Erkenntnis* 48.2, pp. 281–301.
- (2012). "Materialism, dualism, and 'simple' theories of personal identity". In: *Personal Identity. Complex or Simple?* Ed. by G. Gasser and M. Stefan. Cambridge University Press, pp. 206–235.