The Normative Implications of Personal Identity Theory

by

Annette Sonja Dufner

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Philosophy Department
University of Toronto

© Copyright by Annette Dufner 2010
The Normative Implications of Personal Identity Theory

Annette Sonja Dufner

Doctor of Philosophy

Philosophy Department
University of Toronto

2010

Abstract

This dissertation investigates the relationship between personal identity theory and normative ethics. There are a surprising number of contradictory arguments about the relevance of personal identity criteria for ethics and political philosophy. For example, writers like David Gauthier, Thomas Nagel and John Rawls argue that utilitarianism disregards the separateness, or non-identity, of persons. Derek Parfit claims the opposite. According to him, utilitarians respect the separateness of persons while also respecting the disintegration within lives. Also, Henry Sidgwick was under the impression that the separateness of persons supports egoism and disregard for the other-regarding demands of morality. David Brink claims the opposite. According to him, the resulting form of egoism would be nearly indistinguishable from altruism and therefore not at odds with the other-regarding demands of morality. Finally, both Rawls and Nozick argue that the separateness of persons speaks in favour of their own deontological view—despite the fact that their views are fundamentally opposed. Rawls hopes to gain support for his contractually derived principles of justice, while Nozick believes he can generate support for his deontological side constraints. This situation is rather surprising and in need of an explanation. Finally, there have also been skeptical arguments about the force of such appeals, such as Rawls’s claims that personal identity theory and ethics are independent, and his claim that personal identity criteria are inherently normative. This dissertation looks at these controversies in detail and comes to the conclusion that the force of appeals to personal identity is considerably more fine-grained and more complex than has been presumed.
Acknowledgements

I would like to thank a number of people without whom this thesis would be substantially worse. Without my family and especially Daniel Becht this dissertation would not have been possible. Throughout my work on the project, I had helpful discussions with members of the philosophy community at the University of Toronto. In particular, I would like to acknowledge Helga Varden, Shelley Weinberg, Rachel Bryant, Lauren Bialystok and Stephen Thierman who also read near final versions of my chapters. I would also like to acknowledge Jaegwon Kim whose graduate seminar on self-knowledge at Brown University provided some of the inspiration for the choice of this topic. Among the faculty at the University of Toronto, my area committee members Jennifer Whiting, Wayne Sumner and Thomas Hurka helped me to sharpen the focus of the project. Wayne Sumner and Sergio Tenenbaum, my readers, have provided much needed help and advice on various drafts of the chapters. Finally, without the thorough and extensive comments by Thomas Hurka, my supervisor, the chapters would have never evolved into their current shape. I certainly received some of the most excellent, inspiring and encouraging feedback from him that I have had the privilege to receive.
# Table of Contents

1 **Introduction** 2

2 **Personal Identity and What Matters** 19

   2.1 *Self-Concern, Metaphysics and the Normative Realm* 20
       2.1.1 Is Identity What Matters to Us? 20
       2.1.2 Identity and What We Are 22
       2.1.3 Identity and Normative Questions 24

   2.2 *What Matters for Normative Judgments: The Lack of a Further Fact or Psychological Relations?* 28

   2.3 *What Matters for Self-Concern: Psychological Continuity or Connectedness?* 34
       2.3.1 Why Connectedness Might Matter More 37
       2.3.2 Successive and Overlapping Selves 40
       2.3.3 Character Change and Memory Loss 44

   2.4 *First-person Access and Sharing Mental States with Others* 49

3 **Consequentialism and the Separatness of Persons** 53

   3.1 *The Objection against Balancing and Uncompensated Sacrifices* 54
       3.1.1 Nozick on Prohibiting All Involuntary Sacrifices 56
       3.1.2 Rawls on Prohibiting Particular Kinds of Balancing Acts 60

   3.2 *Parfit on Scope and Weight of Distributive Principles* 68
       3.2.1 Parfit’s Ambiguities 76
       3.2.2 Agent-Neutral Quasi-Compensation: A Utilitarian Ideal? 82
       3.2.3 A Holistic Reading of the Early View 88

4 **Egoism, Morality and the Separatness of Persons** 101

   4.1 *Sidgwick: The Separatness of Persons makes Self-Bias Rational* 102

   4.2 *Parfit: The Right View about Survival Undermines Rational Egoism* 104
       4.2.1 Consequence: A Smaller Scope and Less Weight for Self-Bias 105

   4.3 *Brink: The Right View about Persons Supports “Self-extended Egoism”* 111
       4.3.1 Consequence: A Bigger Scope for Self-Bias and Overlap with Morality 114
       4.3.2 Brink’s Omission of First-Person Memories 116
       4.3.3 Brink’s Asymmetry about Discount Rates 122
4.3.4 Egoism, Self-referential Altruism, and Morality 128

4.4 A Hybrid Account about Overlap between Self-Bias and Other-Regarding Concern 133

5 Deontology and the Separateness of Persons 144

5.1 How Separateness might Support Deontological Side Constraints 150
   5.1.1 Taurek and the View that Separateness Speaks against Adding in All Cases 153
   5.1.2 Agent-Centred and Patient-Centred Perspectives on Separateness 158
   5.1.3 Can Separateness Support Side Constraints that Avoid Taurek’s View? 161

5.2 How Separateness Might Support Contractually Derived Basic Rights and Liberties 163
   5.2.1 Contractualism and Respect for Hypothetical Separate Persons 165
   5.2.2 How a Benevolent Spectator Might Also Respect Separateness 167
   5.2.3 Equal Rights and Liberties: Separateness as a Coarse Tool 171

6 What Can Appeals to Personal Identity Establish in Normative Ethics? 173

6.1 The Independence Thesis 174
   6.1.1 Rawls’s Formulation of the Independence Thesis 174
   6.1.2 Two Readings of the Independence Thesis 175
   6.1.3 Independence and Purely Physical Criteria 181
   6.1.4 Independence and Criteria Requiring only Minimal Psychological Content 183
   6.1.5 Independence and Criteria Requiring Diachronic Psychological Content 187

6.2 The Plasticity Thesis 189
   6.2.1 Rawls’s Formulation of the Plasticity Thesis 189
   6.2.2 Non-realist about Personal Identity, Realism about Identity Criteria 192
   6.2.3 Falsifying Plasticity with Conceptual Methods? 194
   6.2.4 Falsifying Conceptual Analysis with Empirical Methods? 195

6.3 Is the Later Rawls at Odds with the Earlier Rawls? 199
   6.3.1 The Rawlsian Appeal to Separateness Re-visited 199
   6.3.2 Independence and Circularity in a Reflective Equilibrium 201

7 Conclusion: A Complex Fine-Grained Underdeterminacy Thesis 203

8 Bibliography 217
The Normative Implications of
Personal Identity Theory
1 Introduction
The last decades of the Twentieth Century have seen an almost incredible array of
contradictory arguments about the relevance of personal identity for questions of justice,
normative ethics and applied ethics. There have been cases in which polar-opposite moral
viewpoints have been supported by appeal to one and the same aspect of personal identity or
vice versa. Even though philosophy is an argumentative discipline, this is rather unusual. The
obvious hypothesis that there must have been some fundamental oversights calls for a
systematic assessment.

One of these controversies arose in 1971 when John Rawls put forth his well-known
argument according to which utilitarianism “does not take seriously the distinction between
persons”\(^1\) and therefore needs to be rejected. Derek Parfit responded in “Later Selves and
Moral Principles”\(^2\) that a proper way of thinking about persons, namely his reductionist way
of doing so, might in fact support utilitarianism.

Without going into any detail for now, here are some more examples: While Rawls
seemed to imply that proper regard for the separateness, or non-identity, of persons would
support egalitarianism, Robert Nozick argued that it would in fact support his fundamentally
opposed libertarian rights view.\(^3\) Long before them, Henry Sidgwick had drawn another,
completely different conclusion: He had argued that the separateness of persons supports
rational egoism.\(^4\) Parfit, on the other hand, was convinced that his reductionist way of
thinking about persons would make the egoist’s self-interest theory less plausible.\(^5\)

\(^{3}\) Nozick, *Anarchy, State, and Utopia*, p. 33.
Brink promptly countered this with the claim that Parfitian reductionism is in fact compatible with the self-interest theory after all.⁶

I argue that such contradictions are due to a lack of detail and subsequent ambiguities about the way in which personal identity matters in ethics. How people think of personal identity can indeed support particular moral principles rather than others, but the support is a lot more fine-grained than has been presumed. Very general appeals to the nature of persons and their persistence across time will establish comparatively little.

This conclusion agrees partly with a restricted reading of the so-called independence thesis, which was first put forth by Rawls in his later writings.⁷ According to the independence thesis appeals to personal identity cannot determine theory selection in ethics. The reason is that proponents of each moral theory can appeal to their own criterion of personal identity that will then lend the most support to their own theory.

I partly agree with the independence thesis in as far as the relationship between personal identity criteria and the choice between entire moral theories such as utilitarianism or justice as fairness is concerned. I argue that appeals to personal identity can lend some relative, though never decisive, support to entire moral theories like utilitarianism or Kantian egalitarianism. One reason for this is that particular criteria of personal identity sometimes only influence the plausibility and the weight of individual aspects of general moral principles, not the entire moral view. I maintain that there are nonetheless some complex dependency relationships between particular criteria of personal identity and particular aspects of moral views, for example the importance a theory will lend to distributive

---

⁶ Brink, “Rational Egoism and the Separateness of Persons,” p. 96: “Parfit’s reductionist claims about personal identity appear to challenge the separateness of persons, I shall consider his claims and argue that the egoist rationale is compatible with reductionism about personal identity.”

⁷ Rawls, “The Independence of Moral Theory.”
principles, the extent of the tension with egoism that a moral theory displays, or the degree of
agent-relativity or agent-neutrality of a moral theory.

There are numerous principles in ethics and political philosophy that require us to
trace their relevant subjects and objects through time. One example is the principle of
equality. If we were to decide that all persons should get an equal share of public resources
over the course of their lives, we would have to be able to identify the relevant persons and
keep track of them through time in order to monitor how much they have already received. If
we lost track of them over time, we might end up giving to the same person twice. Such a
procedure would also require knowing what exactly a person is and subsequently when a
person begins and ends. Otherwise, it would remain unclear whether benefits for embryos or
the extremely demented should count as having contributed to a person’s life total.

We also have to be able to trace persons through time if we want to ascribe various
rights and duties to them, if we want to determine what is in their self-interest or if we want to
grant them autonomy over the various parts of their own existence. Of course such principles
often recommend conflicting actions and measurements. However, they all have in common
that we have to know certain things about persons in order to apply them correctly. We have
to know what the human subjects and objects of these principles essentially are.

1.
What are some of the views on what we are? First of all, there is a commonly held view
according to which our existence is tied to a soul or a metaphysical substance that is at least in
principle independent from our bodies. There are a number of ways in which such a view can
take shape. Often it is thought that we begin to exist at conception and that we might continue
to exist after our bodies have died. If these views are combined with the view of a soul as an
immaterial separately existing entity, then the soul is not something that is necessarily conscious. At conception we have no consciousness yet. Subsequently, consciousness could not be a necessary feature of such a soul. The idea would rather have to be that it is some kind of a spirit that gives our lives value irrespective of whether we are conscious or not. We may call this a featureless conception of the soul.

Alternatively, such an immaterial and separately existing soul could be conscious. According to Descartes “the human soul is always thinking…[It] cannot ever be without thought; it can of course be without this or that thought, but it cannot be without some thought.” His famous correlation of thinking and being in the Meditations seems to capture something that we take to be essential about our existence: our peculiar awareness of our own thinking or consciousness, and the fact that it seems so impossible to doubt it while we experience it.  

However, it is unclear how our special awareness of our own thinking relates to the claim that we are an immaterial separately existing soul. How does our own awareness of our thinking establish that there is an immaterial separately existing soul as the subject of that thinking? How can we be so sure that there is an immaterial separately existing thinker that is doing the thinking when all we are aware of is that “thinking is going on”?  

Another problem with Descartes focus on indubitable consciousness is that he seems to have failed to establish that our soul extends through time. His critics believe that he was not able to show that the thinking soul that exists tomorrow is still the identical thinking soul as the one that exists right now. Arguably, it could be the case that what I experience as “my” consciousness right now is actually tied to multiple consecutive souls. How could we possibly

---

8 Descartes, The Philosophical Writings of Descartes: Volume 2, pp. 189, 336.
know whether the soul that was conscious yesterday and the one that is doing the thinking right now are identical?

If we are really a soul that is in principle separable from the body, and if the relation in which the soul stands to yesterday’s consciousness is not entirely clear, then other strange things would seem to be possible too. As John Locke observed, it would be “possible for Seth, Ismael, Socrates, Pilate, St. Austin, and Caesar Borgia, to be the same man. For if the identity of the soul alone makes the same man; and there be nothing in the nature of matter why the same individual spirit may not be united to different bodies, it will be possible that those men, living in different ages, … may have been the same man.”

Some people might find this exciting rather than problematic, but Locke was decidedly against the idea that we are such an immaterial substance that can be tied to successive conscious humans throughout the ages. One could say that we should reject it for lack of conclusive evidence. Maybe we would have some evidence for it, if persons were regularly able to prove that they had had a previous life in the past, for example by knowing certain details that only a particular dead person could have known. But even this only seems to suggest that they are part of the same stream of consciousness, not that they necessarily have the same soul. So while we have no conclusive evidence against the thesis that we are an immaterial separately existing soul, we also have none in its favour.

Locke wanted to stay away from such speculative concepts. He believed our personal identity consists simply in consciousness that extends through time. According to this view there can only be a person if there is consciousness, and a person’s persistence depends on

---

10 Parfit, Reasons and Persons, section 82.
sameness of consciousness. Locke thought, “as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person…”\textsuperscript{11}

Descartes could object that our temporally extended consciousness such as our memory often deceives us, but Locke nonetheless relies on temporally extended continuity of consciousness for his account of personal identity. On the one hand his view stays away from overly speculative conceptions of the soul, on the other hand he does not push the possible skepticism as far as Descartes did.

Locke also thought that persons seek pleasure, avoid pain and are concerned about their future happiness. Moreover, he understood a person as a “thinking intelligent being, that has reason and reflection”\textsuperscript{12} and that is capable of “appropriating actions and their merit.”\textsuperscript{13} Subsequently, he saw personal identity as important for the forensic purpose of ascribing responsibility for past actions. So for Locke an account of what we are and the use of moral principles such as responsibility were closely tied together.

Identity in the sense in which Locke used the term says that beings at two points in time are numerically identical rather than two qualitatively similar, but actually different beings. Identity is therefore a one-to-one relation. While a thing can be similar to multiple other things, it can only ever be identical to one thing, namely itself.

In recent times, though, neo-Lockeans have raised arguments against the concept of identity that could also have been raised against the featureless conception of a soul. They have argued that it is not the one-to-one identity relation as such that matters, but really only continuity of consciousness. This thought is usually illustrated with Derek Parfit’s

\textsuperscript{13} Locke, “Of Identity and Diversity,” in Perry, \textit{Personal Identity}, p. 50.
hypothetical case of branching. According to such a case, we assume that the two halves of someone’s brain are functional duplicates of each other and get transplanted into the heads of two identical triplet brothers. When the two survivors wake up, they are both psychologically continuous with the being before the operation.

In the case of such psychological branching, the two off-shoots are not identical with the previous being, because identity is a one-to-one relation. Nonetheless, it seems as though what matters to us has been preserved. The basis for the special self-interested concern we have in our own future has remained intact. In a sense it has even been preserved twice. In his book *Reasons and Persons* Parfit concludes what matters about what we are in terms of what we are specially concerned about in a self-interested way is psychological continuity and/or connectedness with any kind of cause.\(^{14}\) It does not matter if the cause for our temporally extended consciousness has to do with technologically advanced scientists who cause this persistence in non-standard ways.

2.

Another subject of controversy when asking what we are concerns the status of our animal bodies. John Locke famously endorsed a dualist account of our personal identity and our animal identity. According to this view, our person and our animal have separate identity conditions. Having one is not strictly speaking necessary to having the other.

It is obvious that there can be a human body without there being a psychologically continuous person at the same time. The best example is a being in an irreversible persistent vegetative state. Also, it might in principle be possible for us to survive the destruction of our

animal bodies as psychologically continuous persons in an afterlife. Locke wanted to remain agnostic about whether there needs to be something physical in order for a person to persist. A central example by Locke that illustrates a further impact of his dualist position is his thought experiment of the body swap between a prince and a cobbler. If the being with the prince’s body behaves like a cobbler after the swap, we would assume that being to be the cobbler—irrespective of its being attached to a body that formerly belonged to the prince. The body’s identity does nothing to determine the identity of the person.

Not all thinkers agree with this separation of personal identity and animal identity. One way of disagreeing with it actually consists in arguing that having a living organism is the only thing that is essential about our existence. Eric Olson puts forth this view in his book *The Human Animal: Personal Identity without Psychology.* According to such a theory, there is only one identity we have and it is our animal identity. There is no separate personal identity in addition to this. Of course human animals sometimes have consciousness, but according to animalism this is merely a temporary property of the animal and nothing that is truly essential about our existence. It is a phase sortal rather than a substance sortal. Human animals do often *not* have any consciousness. For example, there are those that have fallen into an irreversible vegetative state like Terri Schiavo. According to animalism, the essence of Terri was still there when the use of her feeding tube was discontinued, even though she had irreversibly lost all consciousness.

Nonetheless, Eric Olson would not necessarily argue that the machines that kept Terri alive should have kept running. The reason is his belief that moral principles might in fact relate to the property of consciousness that human animals sometimes have. So despite being an animalist Olson believes what matters for moral questions has at least partly to do with

---

15 Compare also DeGrazia, *Human Identity and Bioethics.*
consciousness. There is a difference between what we essentially are and the basis of our moral principles. While we are essentially animals, our moral principles might attach to mere phase sortals of these animals.

It would be strange if Olson insisted on the animal organism as the central basis for all moral judgments. The reason is that there are at least some moral questions that just never arise unless there is consciousness. For example, it does not make very much sense to defend a moral rule according to which one should not offend others, unless there is some kind of consciousness in the picture. At the end of her animal life Terri Schiavo presumably never felt offended nor did she willingly offend anyone else. Such a moral rule would simply not have been applicable to her.

There are also other ways in which thinkers have disagreed with Locke. Locke not only wanted to remain agnostic about whether a person always comes with an animal, he also wanted to remain agnostic about whether consciousness necessarily only occurs when there is something physical, or whether consciousness is in principle separable from everything physical. Contemporary authors tend to believe that consciousness is tied to something physical. They doubt that our consciousness is in principle separable from a physical brain. They also do not tend to believe that we ought to remain agnostic about whether this is possible. Despite taking consciousness to be the central criterion of personal identity, numerous contemporaries believe continuity of something physical such as the brain to be one of the necessary conditions for consciousness and personal identity. This kind of view is different from animalism in the sense that it does not focus on the organism. It merely requires a brain or parts thereof. These parts do not have to be the ones that are important for the functions of the organism such as its circulation or its breathing.
McMahan, for instance, argues in favour of such a double requirement. According to his account, the criterion for personal identity is physical continuity of certain parts of the brain with at least minimal psychological functioning. The parts of the brain he takes to be relevant are those areas of the brain in which the individual’s uniquely personal consciousness is realized. The physical continuity of brain parts is thus necessary for our further persistence as psychological persons. For him, both aspects are essential and necessary conditions of what we are.

There is also another sense in which McMahan’s view is an alternative to other established views. His psychological requirement is also significantly different from the Lockean and the Parfitian views. These views require continuity of consciousness, for example continuity of memory. McMahan’s account on the other hand requires only minimal psychological functioning. The kind of consciousness displayed by late stage Alzheimer patients who have irrevocably lost most memories is perfectly sufficient for him. If there is a centre of consciousness based in a physically continuous brain, this is sufficient for him to say that the person has remained identical. There does not have to be an uninterrupted continuous chain of memories; there merely needs to be a continuously persisting centre of momentary consciousness tied to a physically continuous brain.

3.
There is another question that arises for psychological accounts that do not require a physically identical brain or a human organism in addition to psychological continuity. If we are entirely psychological entities, then the borderlines of what we are might not run exactly along the lines of what we ordinarily assume. Instead, we might be at least partly psychologically continuous with beings that we ordinarily think of as others. After all,
significant others often shape our mindset to a significant extent. Parents and teachers are sometimes thought to have helped turn us into what we are.

This does not mean that all kinds of relevant psychological relations that occur within individual personal lives can also occur within groups of closely related individuals. For example, the unique first-person character of experience memory or the phenomenological character of sense impressions might well be restricted to individuals. Other kinds of psychological relations, though, for example beliefs, desires and intentions as well as the causal ties that contribute to these states might in certain ways occur among significant others as well. The first-person character of recollections and phenomenological sense impressions might occur exclusively within individuals as we ordinarily think of them—even if similarity of beliefs, desires or intentions as well as the causal influences leading to these states are shareable among closely knit personal groups.

If a psychological account of personal identity that does not require identity of brain is true, then we cannot appeal to the fact that there are two separate brains in order to avoid the question whether some psychological relations might not also occur in between individual persons as we ordinarily think of them and not just within such individuals. Even if there is a way of avoiding this possibility, it might still be the case that the psychological relations that hold among significant others are sufficiently similar to the ones that hold within one individual in order to warrant some of the same normative implications. Ethicists who want to defend the view that we have stronger moral obligations toward the members of our own group sometimes put this view forward. What counts as such a group might be a family, a nation, a linguistic group or any other kind of group with which people sometimes “identify” themselves. The idea is that certain psychological similarities and causal ties among the members of such groups warrant additional moral obligations and subsequently lesser
obligations toward outsiders. Others have argued that the psychological relations that connect us with others warrant the ascription of self-interested reasons to units of extended selves\textsuperscript{16} or that there can be group persons that deliberate and act in the same way as individual selves.\textsuperscript{17}

On the other hand, Cartesian intuitions about the mind say that there is something like privileged access to particular mental states that does not seem to apply within extended selves or group agents. The access that a particular individual as we ordinarily think of it has to its own mental states seems to be particularly immediate and epistemically secure in a way that the access to mental states of other parts of a group person is not. It seems to be hard to doubt the thought that thinking is going on while one experiences the thinking. There also seems to be a way in which avowals that an individual as we ordinarily think of it makes about his or her own mental states are peculiarly infallible. My own claim that I am feeling strange right now seems to guarantee its own truth in a special way in which the same judgment about someone else does not.

A complete investigation into what we are therefore has to include an analysis about the extent to which we can share relevant psychological states and relations with entities that we would ordinarily think of as separate persons. We cannot simply assume that there is one person per brain, unless we actually endorse continuity of brain or brain areas as a necessary component within a psychological account of what we are. Quite possibly, this will depend on the kinds of mental states we look at. First-person memories and phenomenological sense impressions might not be shareable, while beliefs, desires and intentions might be shareable with significant others in a relevant way.

\textsuperscript{16} Brink, “Rational Egoism and the Separateness of Persons.”
\textsuperscript{17} See for example, Carol Rovane, \textit{The Bounds of Agency: An Essay in Revisionary Metaphysics}. 

14
In the first chapter of this dissertation I provide the ground for an investigation of the polar-opposite claims that have been made by appeal to the nature and persistence conditions of persons. I discuss different identity criteria by appeal to which we can decide whether a particular person has remained numerically the same across time, and what separates him or her from others. In particular, I address stronger and weaker ways of interpreting Parfit’s influential reductionist understanding of personal identity across time: He is ambiguous about the relative weight of psychological continuity, psychological connectedness and momentary experiences. I also address different kinds of diachronic psychological features of persons that matter for practical purposes, for example intentions and recollections. Overall, this chapter is of synoptic rather than argumentative character.

In the second chapter, I investigate the claim according to which utilitarianism disregards the separateness of persons. The claim that utilitarianism disregards the separateness of persons usually relies on a restriction against balancing benefits and burdens between separate lives. According to such a restriction it should not be possible to outbalance disadvantages in one life with advantages in another. I show that there can be various possible interpretations of such a balancing restriction. I also analyze whether Parfit’s response to this charge against utilitarianism actually works. According to Parfit, a reductionist understanding of personal identity does two things to distributive principles: It increases their scope and reduces their weight. Instead of treating persons’ entire lives as the basis for distributive considerations, we have to distribute between life stages the length of which depends on one’s specific interpretation of reductionism. I show that even on the strongest possible reading of Parfit’s reductionism, the weight of distributive principles will nonetheless never decline to
zero. However, I also show that it can reduce their weight at least somewhat. Therefore, a reductionist understanding of personal identity will in fact lend some relative support to utilitarianism in comparison to theories that give absolute weight to distributive principles, such as egalitarianism.

In the third chapter, I look at the notion of egoism. I argue that our criteria of personal identity will determine our conception of egoism and how strongly the principle of self-interest will usually conflict with the non-instrumental, other-regarding demands of morality. Sidgwick’s egoist believes that he stands in a particular relationship with other temporal parts of himself that never applies between separate persons. This gives him reason to be generally self-biased. On the other hand, Brink’s understanding of this relationship challenges Sidgwick’s argument for egoism. According to Brink, the identity criteria that matter for practical considerations are the connectedness of beliefs, of desires and of intentions. These forms of connectedness can apply not only between the various stages of each individual’s life but also between separate persons, for example between close friends. This will lead to an overlap between “self-extended” egoism and morality, he argues. I raise a number of problems with Brink’s position and propose a hybrid account. According to this hybrid account, there is a limited time frame of particularly vivid first-person connectedness that justifies self-bias in the way Sidgwick thought. However, when deciding whether to benefit one’s own remotely connected future self or the closely connected current self of a close friend, both self-bias and morality will have to favour the friend.

In chapter four, I address deontological arguments. I discern how exactly both Rawls and Nozick try to derive support for their view from the separateness of persons. There is some reason to believe that they merely intended a negative argument against common competitors like utilitarianism. But there is also reason to believe they intended positive
arguments for libertarian side-constraints in Nozick’s case and for contractually derived principles of justice in Rawls’s case. I argue that Nozick’s argument in favour of side-constraints faces a dilemma. On the one hand, he can endorse a holistic understanding of the separateness of persons. In this case, his argument only seems to work if he adopts Taurek’s extreme view according to which we never have reason to rescue the larger number. On the other hand, he can give over-proportional weight to the separateness of persons’ capacities as agents rather than patients. In this case, we are lacking an argument for the weight distribution among the separateness of agents and the separateness of patients that can itself derive support from considerations about separateness. I also argue that Rawls’s argument in favour of contractualism requires giving over-proportional weight to the separateness of persons’ capacities as agents rather than patients. A more holistic understanding of the separateness of persons will not necessarily lend support to a contractualist framework. Moreover, while Rawls is right in believing the separateness of persons gives rise to an initial assumption of equality between persons, a mere appeal to separateness on its own is too coarse a tool to tell us how exactly this equality should feature in our moral theory.

In the final chapter I address two skeptical arguments in the form in which they have been presented in Rawls’s later work. One concern is that personal identity criteria cannot in fact lend decisive support to particular moral views after all. The second concern is that the dependency relationship between personal identity criteria and normative principles might actually run the other way around. I argue that the relationship between personal identity criteria and normative theories is more fine-grained than has been presumed by most authors. Arguments from the nature of persons to moral conclusions have to be very specific in formulating the criteria they appeal to and have to take into account that there are multiple criteria to which we can have pre-normative commitments, all of which can receive various
relative amounts of weight. In the end I settle for a complex fine-grained underdeterminacy thesis to describe the relationship between personal identity and normative ethics.
2 Personal Identity and What Matters
2.1 Self-Concern, Metaphysics and the Normative Realm

2.1.1 Is Identity What Matters to Us?

Since the writings of John Locke, talk about selfhood or personhood as having to do with a soul has largely turned into talk about a metaphysical *identity* of persons. While there is a wider meaning of identity according to which people can have cultural and habitual identities\(^\text{18}\), personal identity in this metaphysical sense says that

\[
\text{a person at time } t_1 \text{ and a person at } t_2 \text{ are identical as long as they are psychologically continuous and no branching occurred.}
\]

It is important to point out that numerical, or metaphysical identity does not mean the kind of qualitative sameness that can hold between numerous things at once. Monozygotic twins, for example, might qualitatively look the same, but they can never be numerically the same. A thing can only be numerically identical to itself. This is the reason for the no-branching requirement.

The second central notion here, next to identity, is psychological continuity. What identity theorists have in mind when speaking of psychological continuity is usually an overlapping chain of psychological connections. Thomas Reid illustrated the difference between the two notions in his well-known example of a general. He imagined a man who robbed an orchard as a boy, later became a soldier who once took a standard from the enemy in battle, and finally became a general. Reid imagined further that the general remembers what he did as a soldier, but not what he did as a boy. As a soldier, however, he still

\(^{\text{18}}\) Schechtman, *The Constitution of Selves*. 
remembered what he did as a little boy. There are psychological connections between the general and the soldier, as well as between the soldier and the boy, but not between the general and the boy. Nonetheless, these connections overlap. They form an uninterrupted chain. Such an uninterrupted chain of connections is thought of as a continuous chain.

Since the writings of Parfit it is now common to believe that the metaphysical fact of identity is not what matters to us about our existence. His well-known illustration of this point consists in the thought experiment of a person who branches. In this thought experiment we imagine a set of triplet brothers, two of which are brain dead. Next, we imagine that the halves of the third brother’s brain are functional duplicates of each other and that they get transplanted into the bodies of his two brain-dead brothers. After the operation the two survivors not only look the same, but also have the same memories and the same intentions for the future as the first brother. Clearly, the two persons after the operation would not be numerically identical to the first brother, because there are two of them. The identity of the original person has ended.

Nonetheless, Parfit argues, the basis of special self-concern has been preserved. What matters to us about our existence is not really whether we have remained identical, but rather the relations of psychological continuity and connectedness between the present and the past or the future. Moreover, it does not matter whether the cause for the persistence of our psychological life is the kind of cause that is present in normal cases. If an advanced scientist causes this persistence externally by using some radical new means such as brain splitting, what matters to us has nonetheless been preserved. Cases of branching that imply the loss of identity while psychological continuity and connectedness are preserved are just as good as ordinary survival, Parfit argues.

19 Reid, “OF Mr. Locke’s Account of Our Personal Identity,” in Perry, Personal Identity, p. 114.
2.1.2 Identity and What We Are

Parfit does not just believe that identity is not the basis of what we care about. He also believes that there is no irreducible metaphysical fact of identity that determines whether a particular being is me or someone else, and that determines whether a particular person still exists. All we have in order to make such judgments are the diachronic psychological relations—which are in principle a matter of degree rather than a function that will always give us a determinate answer as to whether a particular being is me or someone else or whether a particular person still exists. Parfit admits that we tend to talk as if identity plays some kind of a role, but (at least initially) he also insists that we could actually give a completely impersonal description of the world that refers only to the underlying psychological relations.

Parfit’s central argument to this effect, apart from the branching case, consists in his thought experiment *The Combined Spectrum*. The *Combined Spectrum* asks us to imagine a whole range of possible cases. At one end of the spectrum a mad scientist replaces a few of my memories with the memories of Napoleon and a few of my body cells with replacement cells. At the far end of the spectrum the scientist replaces all of my memories with Napoleon’s memories and all of my body cells with replacements. The point of the thought experiment is to show that in the cases in the middle of the spectrum there is no determinate answer as to whether I have remained identical or have died. This, so the argument goes, is evidence against the claim that an anti-reductionist understanding of identity is true, or at least that there is no evidence in favour of it.

The argument concludes from vagueness to the implausibility of an anti-reductionist understanding of identity, or at least to the claim that there is evidence against it and none in

---

its favour. Moreover, the charge goes, anti-reductionism about identity cannot provide the kind of answer that we want in this case. It can only tell us that we cannot tell whether our identity was preserved. The more detailed answer about exact degrees will inevitably be reductionist.

This indeterminacy argument has been of great appeal to many thinkers. However, the less reductionist minded will deny that it works. They will respond that the indeterminacy argument confuses an ontological point with an epistemological one: The fact that we can sometimes not tell whether something is there or not, does not mean that it cannot exist. When applying an ontological concept to real life cases, we have to decide about certain symptoms that will signify whether the conditions for the ontological concept are fulfilled. The fact that we can sometimes not be certain about the meaning of the symptoms does not have to mean that the ontological concept is problematic.

The proponents of a more anti-reductionist understanding of identity therefore believe that the indeterminacy argument is much less interesting and convincing than it might seem. They believe the indeterminacy argument does not actually show that a more anti-reductionist view cannot account for matters of degree. When it comes down to deciding about the specific symptoms that need to be present in order to satisfy the continuity requirement, both kinds of theories would be looking at the same signs. A more reductionist theory would be able to say the person has continued to a particular reduced degree, while the proponent of an anti-reductionist understanding of identity would have to say that the symptoms leave it vague to a particular degree whether the person is still there. This would lead to the conclusion that the more reductionist view can account for the matter more directly, while the more anti-reductionist answer would be somewhat indirect, but nonetheless able to account for degrees.
Reductionists will insist that in the middle cases of the spectrum there is no evidence for whether I have persisted or died. This question will only have an answer if we believe in identity as a substantial further fact that cannot be fully reduced to psychological (and physical) properties and the relations between them. They will also insist that there is no evidence for the existence of such a fact so that we should be content with a reductionist understanding of identity. At the same time, even if we decided that there is such a metaphysical fact, this fact would not help us to answer the question whether the persons in the middle of the spectrum have survived. So at least, we would have further established that identity does not matter.

2.1.3 Identity and Normative Questions
A further question to be asked is whether an anti-reductionist further fact of identity has been given normative force in certain moral judgments. Parfit is of the opinion that there is no such fact. Consequently, our normative judgments about some matters might have to change once we come to accept his view.

Parfit is of the opinion that a reductionist understanding of personal identity will affect what is rational for persons. The psychological relations that matter in a reductionist account are in principle a matter of degree. This supports a discount rate for self-interest on the basis of one’s specific degree of psychological relatedness between two points in time. Parfit also believes his reductionist account will have an impact on moral principles. Most importantly he imagines an impact on distributive principles. He also expects an impact on the desert and the responsibility principles, on questions of autonomy and paternalism, and on issues arising at the beginning and at the end of life. The chapter on egoism will be dealing extensively with rationality considerations and the chapter on consequentialism will discuss the impact on distributive principles. To avoid repetition I will raise a case about the end of life here to
illustrate the potential impact that becoming a reductionist might have on normative judgments.

Of course the truth of reductionism will only change normative views if an anti-reductionist understanding of identity actually had a normative impact in the past.\textsuperscript{21} To illustrate how this might have been the case, imagine the case of a clinical ethicist who has to decide whether the life support machine of a particular human being may be switched off.\textsuperscript{22} This ethicist will be interested in whether there are any psychological experiences left in the being, for example a string of synchronic experiences of pain and pleasure or other sense perceptions. The ethicist will also be interested in whether any diachronic comprehension is left. Most importantly, the ethicist might wonder whether the being is still identical to the competent person who was walking around and living a normal life in the past or whether that person has ceased to exist.

Proponents of the existence of a further fact of identity would approve of the latter question. They would argue the clinical ethicist should be interested in whether the being’s identity has been preserved. If the being’s identity has been preserved, we might have a particularly strong reason to keep it alive. They might believe a willingness to draw a cut-off line for life support at all is due to the fact that we care about an all-or-nothing relation like identity. After all, asking the moral question whether we should switch off a life support system is an all-or-nothing issue.

On the other hand, reductionists about the further fact will insist that the diachronic psychological persistence of the being is essentially a matter of degree. They would say that

\textsuperscript{21} For a view that denies this, see Johnston’s “minimalism.” Johnston, “Human Beings.”
\textsuperscript{22} This is not an example that Parfit himself uses in the context of his very brief discussion of the beginning and the end of life. He considers a related example in his discussion of balancing benefits and burdens across separate people’s lives (Parfit, Reasons and Persons, section 115), which I will discuss in the chapter on consequentialism.
the specific degree of psychological relatedness that the being still displays is the real reason to switch off the machines or to keep them running. We are sometimes willing to make moral all-or-nothing decisions on the basis of a specific degree of relatedness—not via the detour of speculations about the persistence of a metaphysical fact.

This controversy seems to lead to a stalemate. The one side will keep insisting that the normative force comes from our convictions about the continued existence of the fact of identity; the other side will maintain that it stems from the specific degree of psychological relatedness. This leads us back to the point that reductionists can at most maintain that their answer is more direct and does not include a detour via a metaphysical concept.

But there might be a further consideration to be made on behalf of the reductionists. They could argue as follows. Even if identity matters for normative judgments, it will not matter *in addition* to the degree of psychological relatedness. Consequently, psychological relatedness can offer all the needed moral answers on its own, while identity considerations would have to resort to considerations of degree to give those answers. There is nothing about the moral decision that *only* identity can explain, they will insist.

To illustrate how identity might or might not have normative force *in addition* to psychological relatedness imagine the following case:

A committee of ethicists has to decide whether to switch off a life-support machine. The psychological continuity of the patient reaches 6 on a scale of up to 10. They decide that this is a sufficient degree of psychological persistence to keep the machines running. Now they are puzzling over the next question: Does the likely fact that the person’s identity has been preserved possess any moral force over and beyond the normative force of the specific degree of psychological persistence? The relatives
of the patient could insist that it will likely still be *him* who would continue living. Do they have a stronger moral obligation toward the patient as a result of this?

It seems rather questionable whether and how personal identity should matter here in addition to psychological persistence. There seems to be at most one way of answering the question in the positive. Imagine the country where this takes place is extremely poor and there is another patient competing for resources. For economic reasons only one patient can be saved. Both patients will die if their machines are switched off and both of them cannot regain any of their lost psychological capacities. Another ethics committee has decided that the moral entitlement of the competing patient reaches 8.5 on a scale of up to 10. If the ethicists base the rating for their own patient on the degree of psychological persistence, they would have to give him a 6. If they base it on the fact that the person’s identity has most likely been preserved, they might have to give him a 10 though. If both issues matter equally, they would have to give an 8. Only if identity is more relevant than psychological continuity will they be able to give more than the 8.5 points that the competing patient got. This would make all the difference for their patient.

If continued identity has any extra normative force beyond the degree of psychological persistence in such a case, it can at the most strengthen the claims of the being in the hospital bed. It would be implausible to believe it would lower the patient’s claims. The alternative would be to say that identity neither strengthens nor weakens the patient’s claims after we have considered the matters of degree. This would amount to the view that identity does not have an impact on ethical issues. Once we become reductionists about a further fact of personal identity, we will consequently have to change some of our moral judgments if we based them on identity before.
Parfit thinks most people tend to believe in anti-reductionism about their existence at first. They believe there has to be at all times a determinate fact of the matter as to whether they are still alive. Once they come to realize the truth, they will start to give certain normative considerations less weight than before, he believes. The reason is that the lack of a further metaphysical fact of identity, as well as the fact that only psychological relations seem to matter to us when it comes to our survival, might reduce the weight of these moral considerations.

2.2 What Matters for Normative Judgments: The Lack of a Further Fact or Psychological Relations?

In the following I am going to address two ambiguities in Parfit’s view. The first ambiguity lies between the following two claims: (a) what has impact on normative judgments is the lack of a metaphysical fact and (b) what has impact on normative judgments are our diachronic psychological relations as well as the fact that these are in principle a matter of degree. In *Reasons and Persons* he is undecided and points out that insisting only on (a) can lead to extreme moral views. In his later article “Comments” he favours (a). If (b) is true and psychological relations are decisive to a significant extent, then the effects in the normative realm will be somewhat less extreme.

The second ambiguity in Parfit’s account lies between these two claims: (c) psychological continuity and connectedness matter equally to us, and (d) psychological connectedness matters more to us than psychological continuity. Psychological continuity and psychological connectedness are two different concepts and can therefore clearly lead to different results in terms of survival. The continuity requirement is satisfied when there is an uninterrupted chain of connections. This means for example that there can be continuity in the absence of direct connections, as in the relationship between Reid’s general and the young
boy. Their relationship satisfies the continuity requirement, but not the connectedness requirement. Of course continuity always requires some connections. Without some connections there could not be a continuous chain of connections. However, there can be continuity between an entity at $t_1$ and an entity at $t_2$ even though there are no direct connections between those two entities, as in the case of the general and the little boy. I will discuss this second ambiguity in the section about the difference between continuity and connectedness further down.

I will now take a closer look at the first ambiguity. The first ambiguity lies between the claim that the basis of the revision of our normative judgments should be the lack of a metaphysical fact, and the claim that it should be our psychological relations, which are in principle a matter of degree. Consider the implications of the first claim: If there is no metaphysical fact that unites chains of experiences into the experiences of one particular self, then the metaphysical unity of the self is lost. This means that all we could appeal to for our normative judgments would be a succession of more or less isolated momentary experiences.

The second claim, according to which continuity and connectedness are what we should appeal to in our normative judgments, says something else though: It says that our normative judgments should apply to the time periods over which we are psychologically continuous and connected to a relevant degree. Arguably, such psychological units can last fairly long. Someone who has a photographic memory and very clear intentions for the future might remain strongly internally related for several decades.

If the fundamental metaphysical disunity of all parts of our existence is what should have the decisive normative impact, the consequences could be rather severe. For example, in *Reasons and Persons* Parfit describes the following erosive impact on the desert principle:
On the Extreme Claim, since the Reductionist View is true, no one ever deserves to be punished.

He is unsure, though, whether this is true and argues:

…though this Extreme Claim is defensible, it can be defensibly denied.\(^{23}\)

In his more recent article “Comments,”\(^{24}\) Parfit’s position shifts toward the extreme claim. He abandons the ambiguity and moves toward the view that the lack of the further metaphysical fact of identity and the subsequent complete metaphysical and temporal disunity of the self should be decisive:

We would believe that, in the absence of personal identity, these psychological connections cannot carry with them desert or guilt.\(^{25}\)

In a footnote, he adds a comment about his earlier view in *Reasons and Persons*:

I should not have written that we could defensibly deny these conclusions (pp. 325, 343). I do not know whether this is possible.\(^{26}\)

---


\(^{26}\) Parfit, “Comments,” p. 843.
It is worth taking a look at the argument in favour of the view that the lack of a metaphysical further fact is a normatively decisive issue that will lead to extreme normative revisions. The argument essentially consists of two thought experiments combining the branching case with the issue of compensation. One of them addresses the impact on distributive justice\textsuperscript{27}, the other on responsibility\textsuperscript{28}. Both thought experiments have the same structure, so it should be sufficient to look at one of them. Here I will look at the thought experiment about responsibility. A detailed look at arguments about distributive justice will be part of the chapter on consequentialism.

The thought experiment assumes that Parfit branches. His replica is called Backup. We further assume that Parfit committed a crime before he branched. Even though Backup receives a warning and is told to escape, he gets caught and convicted. The authorities argue that Backup deserves a life sentence for Parfit’s crime. Due to the relations of strong psychological continuity and connectedness between him and Backup he stands in the right kind of relationship to Parfit to be punished, they argue. Backup is outraged. The psychological relations are irrelevant, he counters. He did not choose to resemble Parfit and to have his memories, he insists, and he does not agree he can be punished for something that happened before he even came into existence.

Parfit assumes that most people’s intuitions would side with Backup here. If these intuitions are correct, then psychological relations cannot be of practical significance here. This means we implicitly believe that only identity can carry with it desert for punishment. Since there is no identity between Parfit and Backup, we have support for our intuition that Backup should not be punished for something Parfit did before he branched.

\textsuperscript{27} Parfit, “Comments,” p. 840.
\textsuperscript{28} Parfit, “Comments,” p. 838.
The next step of the argument goes as follows. On a reductionist view, identity consists merely in psychological relations and there is no further fact above and beyond them. Even in cases that do not involve branching there is nothing above and beyond particular psychological relations. Since these relations do not carry with them desert or guilt, nobody ever deserves to be punished for a past crime.

Of course this argument will not seem convincing to everybody. Backup’s outcry that he did not choose to resemble Parfit constitutes an appeal to something that the competing identity view can account for: the importance of a particular kind of autonomous or immanent causation. Parfit’s view about what matters says that any kind of cause, including severe medical brain interference, is permissible. So the thought experiment might simply lend extra support to the view that identity, or at least a particular kind of immanent causation, is relevant for moral judgments after all. The reason we should not punish Backup would then be the lack of identity between him and Parfit, which amounts to the view that identity is relevant for normative judgments after all.

Moreover, the argument can also be denied on Parfitian grounds. Just as Parfit argues that psychological continuity and connectedness in the absence of identity is as good as ordinary survival, one may object here that punishment in the absence of identity is just as good as ordinary punishment. In other words, our intuitions may simply side with the authorities and we may believe that psychological connectedness can serve as a basis for punishment here. On this view Backup is simply trying to find a cheap excuse. Maybe Parfit was even counting on our hesitation to punish Backup and decided to undergo the branching operation simply to avoid his punishment.

\[29\] Immanent causation might in turn presuppose identity.
Given these reasons, someone could conclude that the appeal to the thought experiment does not represent an argument at all, but merely illustrates that we are dealing with three possible options: the view that our moral judgments should apply to identity, the view that they should apply to units of psychological relatedness and the view that they cannot apply to any diachronic relation due to the lack of a further metaphysical fact of identity.

In defence of the thought experiment, though, it may be said that it illustrates a little more than just the existence of the three options. It illustrates that the view according to which the lack of a further fact of identity will have revisionary impact on our moral judgments can account for some of the intuitions that an identity ethicist also has: for example the intuition that Backup should not be punished on the basis of his involuntary connection with Parfit. This agreement between the two rival approaches might lend the “extreme view” some plausibility. Not only is its implication reasonable on the basis of its own premises, it is actually also reasonable on the basis of something an identity ethicist would agree with, namely, the conviction that psychological relations cannot be carriers of desert and punishment in the absence of identity.

Nonetheless, many people will remain hesitant to accept the revisionary consequences of such a view.\textsuperscript{30} This is why it seems relevant to point out a further option that might be the most plausible of all. It might be the case that both the lack of a further fact and psychological relations should play a role in our normative judgments. If such a combined approach is adopted, the consequences will be easier to accept. The lack of the metaphysical fact will reduce the plausibility of desert and responsibility ascriptions across the board. At the same

\textsuperscript{30} For a view that accepts the idea that identity does not matter, but also insists this will not have any revisionary consequences, see Johnston’s “minimalism.” Johnston, “Human Beings.”
time, they would retain some of their plausibility within units of strong psychological relatedness. The loss of plausibility might support moral theories that give less importance to these principles, for example utilitarianism. At the same time the retained plausibility within units of strong psychological relatedness would prevent the account from becoming overly revisionary.

2.3 What Matters for Self-Concern: Psychological Continuity or Connectedness?

The second ambiguity in Parfitian reductionism lies between the notions of continuity and connectedness. As the abovementioned example of Reid’s general illustrated, continuity is present if there is an uninterrupted chain of connections.\textsuperscript{31} Since continuity is defined in terms of connectedness, the latter is explanatorily a more fundamental notion. There could be psychological connections in the absence of continuity, but there could not be continuity in the absence of all connectedness. However, there can be continuity between two temporal stages of a being in the absence of connections between \textit{those} two stages.

In \textit{Reasons and Persons}, Parfit leaves room both for the view that continuity and connectedness matter equally as well as for the view that connectedness matters more. It has also been argued by David Brink that only continuity matters. I will look at these suggestions in order. First, it is important to look at Parfit’s position on the relative importance of connectedness and continuity in some detail.

---

\textsuperscript{31} Note that Parfit’s definition of the difference has an additional component. As he argues, continuity requires the presence of strong psychological connectedness. Strong psychological connectedness is present if “the number of direct connections, over any day, is at least half the number that hold, over every day, in the lives of nearly every actual person” (Parfit, \textit{Reasons and Persons}, p. 206). This definition has been picked up for instance by Jeff McMahan in his book \textit{The Ethics of Killing}. His argument in favour of treating continuity as a matter of degree instead of an all-or-nothing relation rests partly on this second understanding of continuity. Of course these two different conceptions of continuity may not always lead to the same results with regard to particular cases.
In the sections on personal identity in *Reasons and Persons* Parfit writes about continuity and connectedness:

I believe that both relations matter. Others may believe that one matters more than the other. But I know of no argument for such a belief. I shall assume that neither relation matters more than the other. (This is not the assumption that their importance is exactly equal. Such a question could not have an exact answer.)\(^{32}\)

Next, he argues against the view that *only* continuity matters. Connectedness also matters, he insists. The central reason he gives for this is the fact that losses of it are met with regret. In saying this, he does not intend an argument to the effect that connectedness matters more. He merely tries to give a reason why it matters to us at all. The reason he sees is the fact that we strongly dislike the idea of losing all memories of our past experiences. When losing all our memories of the past, there can still be continuity as long as there is an overlapping chain of connections left. But this does not seem enough to us. We would strongly regret the loss of all our memories. Only connectedness can make sense of this resentment, because forgetting does not necessarily lead to a loss of continuity.\(^{33}\)

Also, we generally regret the loss of present desires and intentions or our love for particular people, he argues. Given that we genuinely care about our desires, intentions and loved ones right now, we would have to regret the loss of these attachments. Again, we could lose those things without any impact on our continuity. Only connectedness can make sense of these appreciations. And finally, he argues, the same is true of our character. Most people


value certain aspects of their character. They will not want those aspects to change. Only the notion of connectedness can account for this sentiment.\textsuperscript{34}

Even though the quoted passage says that continuity and connectedness are roughly, though maybe not exactly, equally important, Parfit is unfortunately less explicit about his reasons for believing that continuity matters. I will discuss an argument that has been launched in favour of continuity further below.

Despite what the above passage says, there are other sections of \textit{Reasons and Persons} saying connectedness is in principle more important than continuity. For example:

\textit{Psychological connectedness} is the holding of particular direct psychological connections.

\textit{Psychological continuity} is the holding of overlapping chains of strong connectedness.

Of these two general relations, connectedness is more important in both theory and practice.\textsuperscript{35}

Unfortunately, this section does not offer a very explicit reason for this judgment. Given the context, it may be assumed that what is being pointed out is that connectedness is explanatorily prior to continuity. There need to be some connections before there can be any continuity. Presumably, this is what is meant by the claim that connectedness is more important “in theory.”

We may still wonder what is meant by the remark that connectedness is more important “in practice.” Presumably he means it will be important in terms of what matters to

\textsuperscript{34} Parfit, \textit{Reasons and Persons}, p. 301.
\textsuperscript{35} Parfit, \textit{Reasons and Persons}, p. 206.
us about our survival. If this is what he means, though, the passage seems to contradict the
earlier one suggesting that continuity and connectedness matter equally. This means Parfit is
ambiguous on the relative importance of continuity and connectedness in terms of what
matters to us in survival.

2.3.1 Why Connectedness Might Matter More

It is possible to settle Parfit’s first ambiguity with a combined interpretation according to
which both the lack of a metaphysical further fact and the diachronic psychological relations
matter. This route does not seem to be open with regard to the second ambiguity. It does not
seem possible to combine the view according to which continuity and connectedness matter
equally with the view according to which connectedness matters more. The second ambiguity
seems to require a decision in favour of one option and against the other.

There is some reason to believe that Parfit must have thought connectedness matters
more.\textsuperscript{36} In the context of the passage claiming connectedness to be more important he insists
that only continuity can be a criterion of personal identity, since only continuity is transitive.\textsuperscript{37}
Since he also believes identity does not matter for self-concern, though, this leaves some
room for the hypothesis that continuity might be important to generate a transitive criterion of
identity rather than being very important for self-concern.

Identity is by definition a transitive notion. If A is identical to B, and B is identical to
C, then A also needs to be identical to C. Connectedness relations cannot guarantee such
transitivity. If A is connected to B and B is connected to C, it does \textit{not} follow that A is
connected to C. In order to get a transitive identity criterion we therefore have to adopt the
notion of continuity.

\textsuperscript{36} This view that connectedness is more important is also endorsed by D. Shoemaker, “Utilitarianism and
Personal Identity;” “Selves and Moral Units;” “Disintegrated Persons and Distributive Principles.”
\textsuperscript{37} Parfit, \textit{Reasons and Persons}, p. 206.
If the need to generate transitivity is the motivation behind taking continuity to be important, then we are not dealing with a reason that has to do with self-concern. We are dealing with a reason that has to do with identity. This seems to give some support to the understanding that continuity might not matter that much for self-concern in the end and is in the picture for different reasons.

This view would also be in line with a more nuanced definition of continuity in terms of “strong connectedness” that Parfit gives in this section. The definition reads as follows:

*Psychological continuity* is the holding of overlapping chains of strong connectedness.\(^{38}\)

Then he specifies:

Since connectedness is a matter of degree we cannot plausibly define precisely what counts as enough. But we can claim that there is enough connectedness if the number of direct connections over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person. When there are enough connections, there is what I call strong connectedness.\(^{39}\)

These passages sound like a more nuanced way of explaining continuity than the example of Reid’s general. They seem to suggest that continuity, and the identity relation that is based on continuity, has to be understood as a more or less arbitrary way of drawing a line in an


otherwise blurry picture of an ordinarily large number of connections with various temporal durations. In fact, there might actually be “several thousand” per day, he adds.

The reason we have to draw a line in order to determine when continuity obtains rather than being content with an inherently blurry picture seems to be the need for a second psychological relation that can satisfy the requirement of transitivity that the notion of identity demands per definition. As just mentioned, if this is the central reason why continuity is important, then the reason has nothing to do with self-concern. Continuity would be important for transitivity, but not necessarily for self-concern. Our self-concern might well be more related to our exact degree of connectedness instead.

There might be another reason to believe that connectedness should matter more for self-concern. This reason is best illustrated by use of an example. Imagine we ask a young person how much she cares about her distant future self. We present her with two scenarios. In one scenario there are a few less than half as many connections as in ordinary cases between her now and her distant future self. In the second case, there are a few more than half of what is ordinary.

According to Parfit’s definition, the former scenario would simply represent a case in which a certain number of connections hold, while the latter scenario would represent a case in which there is continuity in addition to connectedness. If continuity matters as much for self-concern as the presence of connectedness, then the young woman would have to be twice as concerned in the second case. Her concern on the basis of connectedness would be related to her specific degree of connectedness. In order for continuity to matter equally, it would have to double that amount of concern.

---

This seems counter-intuitive. After all, there are only a few more connections in the second case. It seems much more plausible to say that her concern should depend mostly on her specific degree of connectedness. Continuity might add something, but surely it should not be able to double the concern. This example would therefore speak in favour of the view that connectedness matters more than continuity. It is a speculative argument though. Parfit’s text seems to be ambiguous about the issue. In the following, I will therefore look at a number of counter-arguments against the hypothesis that connectedness matters more.

2.3.2 Successive and Overlapping Selves

To investigate the relative importance of continuity and connectedness further, it will be helpful to recall Reid’s simpler continuity definition by appeal to the example of the general once more. A theory that takes continuity to be central will come to different conclusions about his example than a theory that takes connectedness to be the most important. A theory that takes continuity to be central would come to the following conclusion: There is a persistent unit of concern that lasts from the soldier all the way to the general. A theory that values connectedness, on the other hand, would conclude: We are dealing with two overlapping units of concern here, one that lasts from the little boy to the soldier and a second one that lasts from the soldier to the general.

The second way of thinking about the case means that two or more overlapping successive selves or concern-units can occupy the same biological body.\(^\text{41}\) This may seem surprising. To many it will not appear as an ordinary way of thinking about people. This might speak against giving overly much importance to connectedness on its own.

However, this worry does not seem to lead very far. It is important to notice that not only the connectedness theorist will sometimes come to this conclusion. There are also cases

\(^{41}\) Radden, *Divided Minds and Successive Selves*. 
in which a continuity theorist would have to decide that two successive selves are occupying the same body. For example, in cases of severe and irreparable amnesia even a continuity theorist would have to resort to such a view. If all mental connections between a present being and his past have been severed, then there is also no continuity anymore. Continuity depends on the presence of at least some connections. In case of sudden and complete amnesia there are two separate continuity chains with no connection in between them. The continuity theorist therefore has to agree that two successive selves can occupy the same body.

On the other hand, there is still a potentially relevant difference between the two views. The connectedness theory, but not the continuity theory, allows for the possibility that successive selves can partly overlap. This is what the connectedness theorist would believe in the case of Reid’s general. The being that lasted from the little boy to the soldier and the being that lasted from the soldier to the general overlap at the time of the soldier. The case of amnesia, on the other hand, in which the continuity theorist would also believe in successive selves, involves no overlap.

David Brink has argued that accepting a theory that allows for overlapping selves or units of concern leads to undesirable arbitrariness, because we would never know which self we are dealing with. He takes such units to be long enough to have some inherent interests, and to be able to perform and receive the benefits of actions. But if such overlapping units are deliberating agents, then whose reasons for action should a present being be concerned about, he asks.42

Consider the following psychological life history that is just like the history of Reid’s general, except a little more detailed:\(^\text{43}\)

Here we are dealing with selves that were connected all the way from \(t_1\) to \(t_7\), from \(t_2\) to \(t_{10}\), and from \(t_4\) to \(t_{12}\). Some time periods, however, are not connected. The temporal stages \(t_1\) and \(t_8\), or \(t_2\) and \(t_{12}\) for example, are lacking a connection. At \(t_8\) the being has forgotten almost everything about \(t_1\).\(^\text{44}\) The worry about overlapping selves is that at \(t_6\), for example, the being does not know whether to deliberate and be concerned about the reasons of Joe 1, Joe 2, or Joe 3, nor do other people know how to identify whose reasons the being is or should be concerned about. Another concern Brink raises is the problem that there does not seem to be a particularly “salient”\(^\text{45}\) place in which to draw the lines between Joe 1, Joe 2, and Joe 3.

Before continuing, a clarifying remark about differences between Brink and Parfit is in order. Deliberative thoughts and reasons are not the only kinds of psychological connections Parfit takes to be important. He is also interested in fear of pain or in backward directed recollections—relations Brink does not take to be particularly important. Nonetheless, Parfit

\(^{43}\) Compare, Brink, “Rational Egoism and the Separateness of Persons,” p. 113.

\(^{44}\) For a theory claiming that a person is a cluster of multiple overlapping selves at every point in time see Lewis, “Survival and Identity.”

would agree that one kind of relevant connection is “that which holds between an intention and the later act in which this intention is carried out.”

A defendant of the idea that selves can overlap would respond to Brink’s concerns with a twofold answer. First, it can be argued that the question of which unit we are dealing with always still has a clear answer. At $t_5$, we are simply dealing with a being that has forward directed connections all the way to $t_{12}$. It has backward directed connections all the way to $t_1$. The being at $t_9$, on the other hand only has backward directed connections to $t_4$. It has forgotten everything about its earliest childhood at $t_1$.

Second, and more importantly, what matters is not which “self” or which Joe we are dealing with, but rather which temporal parts are connected with each other. In fact, Brink’s question about “whose” reasons we are dealing with sounds rather anti-reductionist. In Parfit’s view, the question “with whom” or “with whose” reasons or intentions we are dealing does not matter. Furthermore, it is true that there is no particularly “salient” place to draw a line between the various Joes. Arguably this is not important though. According to a Parfitian connectedness theorist, what matters the most is simply how far the direct connections stretch out.

One might want to add a further thought. The question how we can tell whether a particular being displays certain psychological relations to a large enough extent is not unique to the connectedness theorist. Continuity theorists will also have to deal with such epistemological problems in some cases. In the case of a person with temporary amnesia, for example, it will might not be immediately obvious where the memory gap begins and where it ends. This is not only an epistemological problem for the connectedness theorist, but also for the continuity theorist.

---

We set out by looking at the relationship between Reid’s little boy and the general, a case of continuity in the absence of direct connections. The discussion has brought forth that both a continuity theory and a connectedness theory will sometimes have to come to the conclusion that we are dealing with two successive selves in the same body. Only the connectedness theory will have to claim that these successive beings can overlap. This may be seen as a disadvantage, but it can also be argued that the epistemological problem that arises is something that the continuity theory sometimes faces as well. However, before settling the question whether the connectedness view is a plausible reductionist position, it is worth looking at one more argument.

2.3.3 Character Change and Memory Loss

David Brink has also suggested an argument according to which a theory that cares primarily about connectedness cannot account for the value of character change. Character change presumably involves continuity of character in the absence of connectedness of character. We want to maintain that a future character change, especially a positive change, does not imply a loss of self-concern. If connectedness matters the most, Brink argues, we may have a hard time doing so. He therefore thinks we should only care about continuity, and not about connectedness.47

This argument reveals that Brink has different kinds of connections and a different kind of continuity in mind than Parfit. In the context of this argument in favour of continuity he assumes the relevant kinds of relations to be similarity and causal dependence of character. This is quite different from what Parfit assumes; for him the relevant kinds of relations include backward directed memories and forward directed self-concern. Brink, on the other

hand, is not interested in memory connections. Likewise, he is not genuinely interested in self-concern, or only in as far as it is related to character development.

This is most clearly revealed by looking at an example. Imagine a person who mostly enjoyed drinking alcohol and getting into bar fights in the past, but who has changed for the better by now. This person may be described as follows: He or she has lost the psychological connectedness or similarity with the bad character traits and desires of the past. Because the process of character improvement was gradual, he or she still displays character continuity with the past self. If connectedness of character is all that matters, then basis for self-concern has been lost. Brink would find this solution regrettable. We want to be able to speak of the same individual having undergone an impressive improvement. Apparently, this is only possible if continuity of character matters more than connectedness.

The fact that Brink disregards memory relations is shown by the fact that the alleged problem with connectedness would not arise otherwise. In real-life cases of self-improvement people will usually retain memories of their past selves. Despite having lost the connection to the desire to drink alcohol and get into bar fights, there will still be memories of having desired and done those things in the past. In fact, there will not only be memory connections, but there will probably also be continuity of memory. A loss of connections to past desires and character traits does not imply a loss of memory connections or memory continuity.

Moreover, there is a sense in which it is probably false to say that the person at the later stage is not connected to his or her past desires and character traits anymore. People having undergone tremendous self-improvement will presumably contrast their current accomplishment by appeal to their bad character in the past. They may say such things as “back then I had really terrible habits and I am glad I changed.” This means that even though the actual desire to drink and get into bar fights has been interrupted, on a different level there
is still a very powerful connection. The ongoing desire to be sober has presumably been formed as a result of those past desires. Christine Korsgaard has referred to this kind of relation as “authorial connectedness.”

The idea of authorial connectedness is that a particular present mindset is the causal result of a particular active decision in the past. In regard to character change one could say the current character of the person who improved depends on a particular past decision to change and on the desires behind that decision. This is a causal connection with the past. An authorial connection can be present despite a grave lack of similarity of character at two points in time. At the same time it represents a particularly strong kind of causal connection. The reason is that the desires of the past that caused the current character were desires for the very character that came into existence.

In short, we can say that Brink did not see the equal explanatory force of speaking of a **continuous** desire or character trait, and speaking of a memory or an authorial, causal **connection** with a desire that has itself ceased to exist. An authorial connection or a memory connection back to the ceased desire can likewise explain why the person in question is part of an ongoing unit of concern. By drawing this distinction, Brink’s self-improvement case can be explained without appeal to the notion of continuity. It seems, thus, as if connectedness is not only explanatorily prior to continuity, but also equally capable of explaining the value of character change.

Before concluding, there is a potential objection against the thesis that connectedness of memory or authorial connectedness can account for character change, which we should still consider. The objection is the following: Connectedness may have equal explanatory force when it comes to the loss of character-related features like particular desires and intentions,

---

but not when it comes to the loss of experience memories. The strategy of the argument against the priority of continuity consists in explaining the loss of one kind of connection by appeal to another kind of connection, for example recollections. But if the discontinuity also takes place directly on the level of the recollections, then this same strategy will not be open anymore.

Think for example of someone who has not just changed his or her character, but who also happens to have irrevocably forgotten everything that happened before the change. Maybe a continuity theorist is better able to account for the idea that there might nonetheless be a basis for special self-concern in such a case. According to the continuity theory the forgetting of particular events or times does not necessarily lead to a loss of self-concern. The reason is that there is usually a stage of the self in between the present and the forgotten time that still remembers the forgotten events. This means that there is a continuous chain and that we are justified in having special concern. A connectedness theorist, on the other hand, might have to say that the basis for self-concern is gone. According to a connectedness theorist, a person anticipating such a change does not have reason for special self-concern anymore. It is not clear whether this is how we would approach such a change though.

Connectedness theorists can give a number of possible responses to this objection. First, they can try to argue that we rarely forget things irrevocably. Having only temporarily forgotten something does not have to imply a loss of the memory connection, because it is compatible with saying that there is still a capacity to remember. If presented with the right kinds of reminders, people are usually able to remember just about anything.

---

49 This is the way Locke seems to have understood his memory criterion when he writes: “For whatsoever any substance has thought or done, which I cannot recollect, and by my consciousness make my own thought an action, it will no more belong to me, whether a part of me thought or did it, than if it had been thought or done.
Second, they can say that we may appeal to any other kind of connection in order to be able to claim that there is a basis for special self-concern, for example connectedness between the most central beliefs, desires and intentions. Even though the patient does not remember what happened during the time in question, there is certainly still a connection between his most central beliefs, desires and intentions on that day and afterward. This means even connectedness theorists could argue that there is a basis for special self-concern in cases of irrevocable memory loss, as long as there is any other kind of connectedness they can appeal to. They would not have to appeal to continuity to do this.

Third, and maybe most importantly, they can argue that even the continuity theory will occasionally have to come to the same conclusions as a connectedness theorist. This would be the case whenever the amnesia is instantaneous and involves most mental states at once. If the amnesia is instantaneous, then there is a complete lack of connections between the times before and after the incident. This means that along with all connections, the continuity of all chains of connections has been interrupted. So even a continuity theorist would have to admit that there can be cases in which the basis for self-concern has been lost.

A potential escape route for both the connectedness and the continuity theorist has been presented by Jeff McMahan. In order to handle cases in which various kinds of psychological relations disappear at the same time, he insists that self-concern will also go with physical persistence. This allows him to accommodate the intuition that we would still care if we found out that our late-stage Alzheimer’s self will be suffering. In the case of Alzheimer’s, both memory connections and character connections as well as other

by any other immaterial being anywhere existing.” (“On Identity and Diversity,” in Perry, Personal Identity, p. 49, emphasis added.) The word “can” seems to suggest capability rather than actual (and constant) remembering.
connections disappear at the same time and only a string of more or less momentary consciousness remains. According to McMahan’s account one can appeal to physical persistence with minimal psychological functioning as a proper basis for self-concern in such cases.  

This means that we can come to a preliminary conclusion about the connectedness view: Connectedness theorists are able to defend either judgment about irrevocable memory loss without having to appeal to continuity. If they want to argue that there is a basis for special self-concern despite memory loss, they can usually appeal to other kinds of connectedness, for example connectedness of beliefs and desires. Moreover, connectedness theorists are able to explain the value of character change without having to appeal to continuity, because they can appeal to connectedness of recollections or authorial memory. If both kinds of connections get lost simultaneously, they can consider appealing to a physical basis for self-concern. They can also argue that even continuity theorists will have to accept a loss of all psychological bases for self-concern in some cases. A theory according to which connectedness matters more should therefore be taken particularly seriously as a potential competitor in debates about the appropriate weighting to attribute to connectedness and continuity.

2.4 First-person Access and Sharing Mental States with Others

The previous section has shown that there can actually be different kinds of psychological connections. There is a way in which this fact might matter that has not been addressed yet: Some of these relations might actually connect us psychologically with significant others.

50 McMahan, *The Ethics of Killing*, pp. 67-68. Also see his thought experiment The Suicide Mission, in McMahan, *The Ethics of Killing*, p. 57.
Others again, might usually only occur within individual persons as we ordinarily think of them.

Brink believes the kind of psychological similarity and causation that matters for self-concern also occurs among significant others. More specifically, what he has in mind is similarity of deliberative and character-related mental states such as beliefs, desires and intentions as well as the causal relations that lead to them. As mentioned before, he disregards first-person memory and the first-person character of sense impressions, so that all the kinds of relations he takes to be relevant might also occur among significant others, and not just within individual lives.51

This can have significant normative implications. Ethicists have referred to the shared psychological relations among significant others to defend the idea that we have stronger moral obligations toward the members of our own group. Such a relevant group could be a family unit, a country, a linguistic group or any other group that people tend to “identify” themselves with. If these authors are right, then we might have a stronger obligation to help our own children than the suffering people on the other side of the planet, because we stand in a closer psychological relationship with them.

People who live with each other often come to share values as well as other beliefs or plans for the future. Due to having spent time with each other they also often have a history of shared experiences and activities. They influence each other’s convictions about the kind of life they want to live, which national or international policies to support or simply what to do for fun. People sometimes even say about their parents, teachers or friends that they helped turn them into what they have become.

_____________________

51 Brink, “Rational Egoism, Self, and Others.”
Proponents of the communitarian tradition have argued that our social relations are a significant aspect of what we are. According to such views, these social relations should receive sufficient acknowledgment in the way we construct our moral views. Theories that rely on overly individualistic conceptions of persons’ preferences are rejected as violating this fundamental feature about the existence of persons. The idea is that persons do not enter the social and political deliberation process with pre-made preferences; instead, their preferences are a product of the social process.

Nonetheless, we do not want to say that a social process can make persons close to indistinguishable. Some of their beliefs, desires and intentions will remain incompatible and separate. Also, even very close people generally spend enough time apart in order to have substantial separate experiences. Moreover, there seems to be something rather solitary about the phenomenological features of our first-person experiences. What it feels or even “looks like” to remember something “from the inside” is arguably not shareable in the same way as value systems, political convictions or propositional beliefs about the past.

It is often said that good friends share a lot of memories, but this cannot mean that they remember the exact same spatial perspective on last night’s dinner table. What they share might be what can be communicated linguistically. They seem to share the propositional content of particular beliefs about the past. How exactly the dress of one of the dinner guests looked to one’s colour-blind best friend is simply not accessible. The phenomenological features of experiences and memories might only be accessible to individual persons.

These intuitions say that our access to our own mental states is particularly immediate in a way that our relationship to the mental states of other persons is not. They also say that our knowledge of our own mental states is epistemically particularly secure. It seems hard to doubt the thought that there is thinking going on while we experience that thinking. There
also seems to be a sense in which pronouncements about one’s own mind are in a special way infallible. If I say or think that I am feeling strange right now, then the first-person thinking of that thought seems to guarantee its own truth in a peculiar way. None of these things seem to be true for the extended selves or group agents that I can form with others.

This means we might be dealing with at least two kinds of psychological states. Phenomenological experiences seem to have a special status that restricts them to individual persons. On the other hand, there seems to be a widespread conviction that deliberative thoughts like beliefs, desires and intentions can in principle be shared with others—even though they will also often remain distinct in separate persons.
3 Consequentialism and the Separateness of Persons
It is still a commonly held view that utilitarianism disregards a fundamental metaphysical fact about persons. It has been argued it disregards the fact that persons are fundamentally separate and not identical with each other. In the early 1960s Gauthier wrote in *Practical Reasoning*: Maximizing for mankind “is to suppose that mankind is a super-person.”\(^{52}\) In 1970 Thomas Nagel argued in *The Possibility of Altruism* that utilitarianism “treats the desires...of distinct persons as if they were the desires...of a mass person.”\(^{53}\) In 1971 John Rawls repeated once more: Utilitarianism implies “conflating all persons into one” and “does not take seriously the distinction between persons.”\(^{54}\) In his book *Reasons and Persons* Derek Parfit tried to offer a response to these accusations.

### 3.1 The Objection against Balancing and Uncompensated Sacrifices

In the context of their objections against utilitarianism, these authors tend to object to the idea that benefits or burdens in one person’s life can be morally outweighed by benefits or burdens in a separate person’s life. This objection to balancing is often tied to the thought that there cannot be compensation between separate persons.\(^{55}\)

Robert Nozick’s passage on this reads as follows:

> Individually, we each sometimes choose to undergo some pain or sacrifice for a greater benefit or to avoid a greater harm…. Why not, similarly, hold that some persons have to bear some costs that benefit other persons more? But there is no social

---

55 In addition to the upcoming passages, support for this thesis comes also from Parfit, *Reasons and Persons*, pp. 336-39 as well as from Brink, “The Separateness of Persons, Distributive Norms, and Moral Theory,” p. 255.
entity with a good that undergoes some sacrifice for its own good…. To use a person in this way does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has. He does not get some overbalancing good from his sacrifice, and no one is entitled to force this upon him…."

Later he argues that moral side constraints “reflect the fact that no moral balancing act can take place among us; there is no moral outweighing of one of our lives by others so as to lead to a greater overall social good.”

Rawls uses similar language in his well-known objection to balancing:

Each member of society is thought to have an inviolability founded on justice or, as some say, on natural right, which even the welfare of everyone else cannot override. Justice denies that the loss of freedom for some is made right by a greater good shared by others. The reasoning which balances the gains and losses of different persons as if they were one person is excluded.

To cite one more author, Nagel explains his claim that utilitarianism “ignores the distinction between persons” in the following way: “To sacrifice one individual life for another, or one individual’s happiness for another’s is very different from sacrificing one gratification for

56 Nozick, *Anarchy, State and Utopia*, pp. 32-33. For Nozick’s Closest Continuer Theory of Personal Identity see Nozick, *Philosophical Explanations*, pp. 27-70. The view says in branching cases personal identity persists in that off-shoot that is the closest continuer to the pre-branching person. It leaves unexplained, though, how we should handle cases of exactly symmetrical branching.


another within a single life.”\textsuperscript{60} He goes on to defend this position by appeal to “the extremely strict position that there can be no interpersonal compensation for sacrifice.”\textsuperscript{61} I will now investigate the work the objection is doing for Nozick and Rawls and later turn to Parfit’s answer and defence of utilitarianism.

3.1.1 Nozick on Prohibiting All Involuntary Sacrifices

David Brink argues that the prohibition against balancing might support a Pareto constraint, which would privilege Nozick’s view over Rawls’s. More specifically, he believes this would follow if the objection to balancing and uncompensated sacrifices is to be understood as an objection against all balancing. Such a view would prohibit balancing whenever it “imposes an uncompensated or net loss of welfare—however small—on one person in order to provide benefits—however great—to others.”\textsuperscript{62}

A situation is Pareto superior to another if there is at least one person that would be better off in it than in any alternative scenario and nobody would be better off in the alternative scenario. A situation is Pareto inferior to another if at least one person would be worse off in it and nobody is better off. A situation is Pareto optimal in case there is no scenario that is Pareto superior to it. Finally, a situation is Pareto incomparable to another if at least one person would be better off in it and at least one person would be better off in the alternative scenario. A Pareto side constraint would prohibit all changes that are not Pareto improvements. This means that balancing acts would always be prohibited unless they produce a situation in which at least one person is better off and nobody is worse off.

\textsuperscript{60} Nagel, \textit{The Possibility of Altruism}, p. 138.
\textsuperscript{61} Nagel, \textit{The Possibility of Altruism}, p. 142.
A number of remarks have to be added to this. First, one might not be allowed to use the distributive status quo as the baseline for Pareto comparisons. If the status quo were the baseline, the default would fall back on it whenever there is no superior option. This would seem to privilege Nozick’s view over Rawls’s, because only he can take the status quo to be unproblematic. Nozick believes there is a natural process of acquisition by which people came to own what they have. This process can occur in unobjectionable ways, he believes, so that the result—the status quo—would also be unobjectionable. Rawlsians would have a different view on this. They would argue that what people have in the status quo has not been established as being rightfully theirs yet. They believe that the relevant base line or starting point for Pareto comparisons should be a hypothetical just distribution scenario. So the claim that separateness supports a Pareto side constraint might not be able to settle the question whether Nozick or Rawls has a stronger argument. We first have to decide what the proper baseline is.

In the search for Pareto optimal improvements we also run into another general problem that makes the question about the proper baseline decisive. The problem is that the Pareto strategy will often not yield a solution unless we make further decisions. For example, if several parties compete for limited resources, it can be impossible to find a solution by appeal to Pareto considerations alone. If we give person A what she wants, then person B will get less than what he wants. Likewise, if we give the resources to B, then A will get less than what she wants. The problem is that the two scenarios would then be Pareto incomparable, because one person is better off in the one scenario and another person is better off in the alternative scenario. Moreover, both alternatives would be Pareto optimal, because there is no scenario that would be better. We are thus dealing with multiple Pareto optimal solutions and there is so far no way of deciding between them.
The natural reaction to this would be to demand that people lower their aspirations until one single optimum can be found. But should everyone have to lower his or her expectations equally or should some people, for example the particularly rich, have to lower them more? This is mostly a problem for Rawls. Nozick would presumably be happy to fall back on the status quo whenever a single Pareto optimal improvement cannot be reached.

For now, we may note that an objection against all balancing would either have to take the status quo as its default, or it would have to be supplemented with a procedure or principle for the lowering of expectations. Otherwise, it might not be able to settle distribution questions at all. If we take the status quo as the default, the objection against all uncompensated balancing would support Nozick’s view. The reason is that only he can take the naturally evolved status quo to be unproblematic.

However, we do not know yet whether the status quo can be defended by appeal to the separateness of persons any more than any particular procedure or principle for the lowering of expectations. Can the natural evolution of the status quo be defended with an appeal to the separateness of persons? If the process respected the separateness of persons, then separateness could presumably at least not count against the natural process.

Followers of Nozick will argue that the separateness of persons was carefully respected in the process. They will argue that the initial acquisition of rights as well as the transfers of existing rights up to now respected the separateness of persons. In regard to the rights transfers, they could argue that the transfers respected the restriction against all Pareto inferior balancing as long as all parties have always consented to them. If all parties have always consented to the transfers, libertarians would not see an infringement of the separate ends and experiences of the parties.
At the beginning of the chain of rights transfers lies a Lockean account of initial acquisition through mixing one’s labour with a previously unowned thing. Does initial acquisition respect the restriction against all Pareto inferior balancing? The issue seems to be that during an act of initial acquisition there were presumably no other actual persons around who were mixing their labour with the particular thing. However, there are persons who would have mixed their labour with the thing had they been around and had others not done it. These people are now unable to make use of opportunities they would otherwise have had.

Nozick argues that this concern is taken care of in the form of a Lockean proviso. He believes that there should be limits on the initial acquisition of property as well as the subsequent transfer of property rights “if the position of others no longer at liberty to use the thing is thereby worsened.” Nozick focuses on situations in which someone’s initial acquisition blocks the total supply of a substance necessary for the survival of others. A person “may not appropriate the only water hole in the desert and charge what he will.”

However, on his view the previous status quo of those who would be worse off is decisive. For example, if a researcher finds a substance in a far away place that can save lives and he refuses to sell it, except on his terms, this does not violate the proviso. The reason is that it does not worsen the situation of those who cannot buy the substance. The researcher’s activities did not worsen their chance of survival.

It is beyond the scope of this investigation to assess the workings of the proviso. But there does seem to be a mechanism in Nozick’s work that is supposed to take care of the complaints of actual persons who cannot get what they need, but who could have participated

---

63 Nozick, Anarchy, State, and Utopia, p. 178.
64 Nozick, Anarchy, State, and Utopia, p. 180.
65 Nozick, Anarchy, State, and Utopia, p. 181.
in the thing’s initial acquisition had they been in the right place at the right time and had others not taken it. Overall, Nozick’s restrictions against balancing can be seen as living up to a fairly strong restriction against one-sided involuntary balancing.

3.1.2 Rawls on Prohibiting Particular Kinds of Balancing Acts

If the separateness of persons supports a Pareto side constraint without a special exclusion of the status quo as the proper baseline, and the initial acquisition of property respects the separateness of persons, then Nozick seems to be doing comparatively well and Rawls might have a problem. A Pareto side constraint would prohibit all changes that are not Pareto improvements. Rawls would be able to agree with such a Pareto side constraint only if we take a just and equal distribution as the baseline. On the other hand, he would also have to be in favour of changes away from the status quo and toward equality—changes that might not be Pareto improvements, because they might make the rich worse off. He would not be willing to endorse a Pareto constraint on moving away from what he takes to be an unjust distribution.

Is there a way in which Rawls can keep up? One possibility would be to argue that the separateness of persons supports his position about a just and equal baseline for Pareto improvements. The idea could be that there are separate persons and that there is nothing within any of these separate units that makes any one of them more important than another. Therefore, the separateness of persons supports an initial assumption of equality of some kind between them.

If this is how one looks at the issue, then insisting on respect for separateness just seems to be another way of insisting on an initial equality assumption. It is unclear whether this would be an argument with a premise and a conclusion, rather than two different ways of saying the same thing. Demanding respect for separateness will just mean to demand initial
equality. The working assumption is the same in both cases. It is the idea that there should be an assumption of initial equality of some kind between separate person units.

While this may be a very plausible response to the separateness of persons, it does not in and of itself show that we should, for example, ensure an equal distribution of opportunities, as Rawls believes. In fact, utilitarians may argue that they also endorse an assumption of initial equality. They believe that everybody should count for one and nobody for more than that. So if separateness merely speaks in favour of an initial assumption of distributive equality, this will still not be a particularly decisive argument.

Maybe there is an additional way of thinking about this. One of the main lines of criticism against Nozick’s view is that some might end up with very little while others can end up with very much simply because of having been lucky enough to be in the right place at the right time. By giving weight to the objections of those who could not participate as much as others in initial acquisition one can ensure that everyone gets a fair share irrespective of matters of sheer luck. This concern directs our attention at those who are worse off through no fault of their own.

Rawls takes account of this concern in the form of his difference principle, according to which social inequalities are to be arranged so that they are to the greatest expected benefit of the least advantaged. This principle goes a lot further in its distributive demands than Nozick’s proviso. For example, it still gives priority to the worst off when they are already pretty well off and not simply struggling for access to the only water hole in the desert. Instead, Rawls would favour the worst off in all cases. More importantly, unlike Nozick, he would endorse involuntary one-sided transfers of wealth from the better off to the worse off. In this particular sense, his view does not live up to the prohibition against balancing as much
as Nozick’s. Changes toward a distribution that he takes to be more just would be warranted for him even if they are not Pareto improvements.

Rawls’s way of putting things as a restriction against balancing, according to which it is not permissible to let the burdens to one person get outweighed by benefits to someone else, is consequently somewhat unfortunate. As an egalitarian, Rawls himself has to be in favour of the idea that some involuntary burdens to one person can be outweighed by some benefits to another. If there were a society in which almost everybody were very well off, but a handful of people did not benefit from this wealth in any way and were doing very badly, Rawls would support the imposition of a small involuntary burden on the well off in order to help those who are the worst off.

He would do so by insisting that the well off would have to agree with this behind the veil of ignorance, but in the later, actual society the transfer could still turn out to be involuntary. Rawls’s understanding of the difference between a voluntary and an involuntary balancing act focuses on the kind of hypothetical consent that persons would give behind the veil of ignorance rather than any actual consent (or lack thereof) that they would give under actual circumstances. Arguably, though, the appeal to hypothetical consent under contractualist considerations can still lead to a justification of actual involuntary transfers under the circumstances of the later society.

As Dennis McKerlie points out, Rawls could have meant at the most that unrestricted balancing would be problematic. Rawls would want to restrict it to cases in which one can appeal to hypothetical consent in the original position. Generally, though, if a prohibition of unrestricted balancing is all that the separateness of persons leads to, then the normative force

\[ \text{(66) McKerlie, “Egalitarianism and the Separateness of Persons.”} \]

\[ \text{(67) McKerlie, “Egalitarianism and the Separateness of Persons.”} \]
of the appeal to that separateness may turn out to be fairly broad. More specifically, without a further argument from separateness to contractualism (which I will critically assess in the chapter on deontology) it will not support only Rawls’s appeal to hypothetical consent as a restricting device. Instead it might support any theory that objects to unrestricted involuntary balancing.

Most theories find it necessary to give strong and restrictive reasons for any involuntary balancing act. Nozick objects to unrestricted involuntary balancing. At the most, only the conditions that will activate the Proviso can justify balancing. Also, utilitarians would clearly disapprove of a policy that deprives one half of the population of much of their happiness in order to achieve marginal benefits for the other half. This means an objection to unrestricted balancing would not even rule out utilitarianism, as the context of Nozick’s and Rawls’s objections against balancing suggests. In fact, virtually all moral frameworks pose some restrictions on balancing. A prohibition against unrestricted balancing might even be the very aspect that unites all moral theories that are in any way concerned with the relationship between different people.

In fact, even a principle that is essentially the exact opposite of egalitarianism would pose restrictions on balancing: Think of an elitist maximax principle as appealed to by Nietzsche and according to which we always ought to promote the excellence of the few most excellent and valuable persons in order to maximize the accomplishments of the human species. If we can distribute a given good that will further the excellence of its recipients, such a maximax principle would demand giving everything to the most excellent person in order to produce the highest possible accomplishment of excellence in society. The worst off in terms of excellence and resources should receive nothing. Such a maximax principle is

---

68 Hurka, *Perfectionism*, p. 75.
basically the opposite of Rawls’s view. Nonetheless, it would also object to unrestricted balancing. It would do so on the ground that balancing could amount to taking resources away from the most excellent.

There is a sense in which any kind of actual one-sided involuntary balancing act seems to transgress the separateness of persons in some way. Even a balancing act that takes from the extremely rich to prevent someone else from starving to death seems to transgress the separateness of persons in the sense that it would impose a one-sided uncompensated burden that the rich might not actually consent to. The same even seems to be true for the distribution of fundamental rights and liberties. Respecting the rights of one’s neighbours means abstaining from doing certain things that would interfere with their rights to do certain other things. What we are dealing with here is likewise a potentially one-sided involuntary trade-off.

Apart from focusing on hypothetical consent, which I will discuss under the heading of Deontology, and apart from the belief that separateness supports an initial assumption of some kind of equality, what Rawls had in mind is a restriction against balancing that favours the worst off. Is there a way of constructing a link between separateness and a view that assigns priority to the worst off? If separateness of persons directs our attention toward the separate experiences and opportunities persons are facing, it might somehow draw our attention to a particular extent to those whose circumstances are particularly undesirable. If this is true in some way, then Rawls’s view could be said to pay particular attention to the

---

69 McKerlie seems to agree with this point. In “Egalitarianism and the Separateness of Persons,” p. 210, he argues: “So it is also an application of the objection to say ‘The losses in welfare for some people cannot be morally outweighed by the gain of respecting someone else’s basic liberty.’ There is no obvious path from the objection to balancing to the conclusion that we should respect one person’s freedom by sacrificing gains in welfare for others. The objection cuts both ways… If an example involves gains for some and losses for others the objection to balancing cannot tell us what to do.”
separateness of people’s circumstances in a way in which Nozickians, and possibly utilitarians, do not.

There seem to be at least two ways of constructing such a position. The first is not a view Rawls would endorse, but will serve the purpose of contrast. First, one could argue that the separateness of persons will lead to a particular focus on the worst off due to the diminishing marginal utility of giving to the better off. Second, one could argue that the separateness of persons will lead to a particular focus on the worst off because it supports the moral view that there is over-proportional intrinsic value in giving to the worse off, rather than just instrumental value as the effect of diminishing marginal utility implies.

Diminishing marginal utility means that the utility of a given thing will diminish the better off its recipients already are. For example, the utility of a given portion of food will be greatest if it is given to the most hungry. If we give the same amount of food to someone who is not hungry, utility will be lower. It might diminish entirely or even become negative if we give the food to people who just had huge meals and would need to force themselves to eat it.

Combining a concern for separateness with the fact of diminishing instrumental utility will support giving priority to the worst off. If we can distribute any instrumental goods, such as money, efficiency will demand that we give priority to the worst off. This will make the best use of the resources available and will increase peoples’ benefit the most overall.

A problem for this solution is that the separateness of persons might actually fall out of the picture in the end. What might be doing the work is the efficiency consideration that the diminishing marginal utility consideration recommends rather than the separateness of persons. If giving one lunch to the starving is simply more efficient at maximizing utility than giving one lunch to the satiated, then the efficiency consideration seems to be a sufficient

---

reason for doing so. This view does not necessarily need the additional claim that the separateness of persons speaks in favour of giving to the worst off. In fact, the separateness of persons may speak in favour of treating everyone equally instead. It is the diminishing marginal utility of giving to the better off that would nonetheless favour the worse off. This means the separateness of persons might not be doing the decisive bit of work in this solution.

Alternatively, one should therefore ask whether separateness might support an account that gives *intrinsic* value to benefiting the worst off. According to such a view, assigning priority to the worst off would not just be a result of efficiency considerations, but has additional value in and of itself. For example, take two people, one of them rich and happy, the other one poor and sad. According to diminishing marginal utility giving one unit of an instrumental good to the poor person will produce a bigger increase of pleasure (or some other intrinsic good) for the poor person. The alternative view will agree with that, but will add a further claim on top. According to this further claim improving the poor person’s life by one unit of pleasure makes a bigger moral difference than improving the rich person’s life by one unit of pleasure.

The two views will often yield different results. For example, it might be the case that we have to choose between the following two options. We might be able to give 100$ to the rich person or 1$ to the poor person. Moreover, it might be the case that—after counting in the effect of diminishing marginal utility—the 100$ will produce three units of pleasure for the rich person, while the 1$ will create only one unit of pleasure for the poor person. In this case, the first view would speak in favour of giving to the rich person, because doing so will be most efficient. The second view that assigns intrinsic value to giving to the worse off can decide otherwise, though. Despite the fact that giving to the rich person will create more pleasure, it can nonetheless recommend giving the 1$ to the poor person on the ground that
one unit of pleasure for a poor person is intrinsically more valuable than three units of pleasure for a rich person.

Diminishing marginal utility is a simple fact about most instrumental goods. The more we have of them, the less intrinsic value such as pleasure they give us. The alternative prioritarian view, on the other hand, is a normative view. It holds that in addition to what the diminishing marginal utility recommends, there is intrinsic value in giving to the worse off. This means diminishing marginal utility will tend to favour the worst off when it comes to the distribution of instrumental goods such as money. Prioritarians will also tend to favour the worse off in the distribution of instrumental goods, because they do not deny the effect of diminishing marginal utility. In addition, though, they will also tend to favour the worse off in the distribution of intrinsic goods, such as pleasure or knowledge.

On this view the separateness of persons might not fall out of the picture as easily. The reason is that it is not simply an efficiency consideration that speaks in favour of giving to the worse off. Instead, there is also intrinsic value in doing so. This intrinsic value might have its ground in our experience of ourselves as separate persons with separate sets of experiences. Irrespective of any efficiency considerations we might find it intrinsically worse if some separate persons are doing a lot worse than all others. This means a prioritarian view is able to give some substance to Rawls’s claim that the separateness of persons speaks in favour of his difference principle.

While Nozick seems to respect the objection against balancing to a stronger extent, Rawls’s interpretation of the force of separateness consists in an initial assumption of equality of some kind and a non-instrumental form of prioritarianism. This means we seem to have found a partial point in favour of Rawls. He might actually respect the conditions of separate persons’ lives in a way that Nozick does not.
Interestingly, the claim that intrinsic prioritarianism actually follows from the separateness of persons would also help Rawls in his efforts to launch a negative argument against utilitarianism. Utilitarians do not believe in the prioritarian claim that there is intrinsic value in favouring the worse off. If separateness supports prioritarianism, then Rawls might have a negative argument from separateness against utilitarianism apart from the objection against unrestricted balancing that utilitarians will also endorse in some way.

The remaining problem is that there is a difference between Rawls’s principles of justice and consequentialist prioritarianism. Rawls’s balancing prohibition favours the worst off more absolutely than consequentialist prioritarianism would. Consequentialist considerations will always allow for some aggregation: If we have to choose between a minor loss for the worst off and a major gain for an extremely large number of persons who are doing only marginally better than the worst off, consequentialists can decide in favour of the gain. Rawlsians would have to decide otherwise. David Brink describes this view as a dictatorship of those with the worst complaint.\textsuperscript{71} It remains unclear whether the separateness of persons is a sharp enough tool to differentiate between these alternative views.

### 3.2 Parfit on Scope and Weight of Distributive Principles

Parfit’s answer to the alleged utilitarian disregard for the separateness of persons is based on his reductionist account of personal identity. He agrees with the critics that utilitarians care less about distribution between separate persons than others. But he also believes that caring less about it is reasonable if one holds a reductionist account of personal identity.

His reductionist account of personal identity attacks the idea of a metaphysical “further fact”\textsuperscript{72} of identity that explains why we should view the various temporal stages of a person’s life as stages of one and the same person. He finds this concept questionable and argues it should not matter in our practical deliberations. What matters instead when we speak about persons can be reduced to experiences at particular points in time and the relations between those experiences. Moreover, these relations are a matter of degree. It is not the case that two particular experiences are either simply connected or simply disconnected. They can be connected to various degrees. Parfit’s account relies on the belief that persons are to a significant extent temporally disintegrated. As he puts it in \textit{Reasons and Persons}: “A Reductionist is more likely to regard this child’s relation to his adult self as being like a relation to a different person.”\textsuperscript{73}

An extremely anti-reductionist position would stipulate that the metaphysical fact of personal identity is something like a separately existing soul. Contemporary versions of anti-reductionism do not refer to a soul anymore, but they believe the metaphysical fact of personal identity as an explanatory entity is doing work that cannot be done by anything it could get reduced to.\textsuperscript{74} Parfit argues that the notions of psychological continuity and connectedness can do all the explanatory work and we need not believe in any metaphysical fact above and beyond them. He acknowledges that there \textit{might} be such a metaphysical fact and that we talk as if there were one, but he also believes that we do not have enough evidence for it, that it is not what matters to us in survival, and that it should therefore also not matter in our practical deliberations.

\textsuperscript{72} Parfit, \textit{Reasons and Persons}, part III.
\textsuperscript{73} Parfit, \textit{Reasons and Persons}, p. 335.
\textsuperscript{74} For example S. Shoemaker, “Self and Substance,” p. 284.
Parfit argues that reductionism does two things to the debate about egalitarian distributive principles: First, he believes it supports the idea that the scope of distributive principles becomes bigger. And second, he argues that the overall weight of distributive principles becomes smaller. As a net result we would have to care less about the distribution between separate people. This, he hopes, could lend support to the comparatively distribution insensitive utilitarian maximization principle.

The first point says that distributive principles do not only apply between different people, but also between the different parts of individual lives. The relevant units for distribution are not simply people any more, but also the various parts of a single person’s life. Instead of “Joe” as the unit for distribution there are actually “Joe as a child”, “Joe as an adolescent” and so forth. While the size of the relevant units becomes smaller in this way, their number and, as a result, the scope for distributive principles becomes bigger. To illustrate this, imagine that someone had to choose between the following two options.

(A) giving Smith 10 benefits at the age of 40, or

(B) giving Smith 5 benefits at the age of 40 and 5 at the age of 50.

It is assumed the benefits are all the same.

To keep the focus of the investigation manageable, we have to leave aside issues about competing theories of the good or holistic goods that are not tied to particular points in time.

If the distribution between the temporal parts of one’s life is important, as Parfit suggests, then scenario B would have to be preferred. In scenario B it is not only a single stage of Smith that benefits. Instead, the benefit is spread out equally between the forty year old and the fifty year old. If the separateness of persons as construed by the mentioned critics of utilitarianism
is more important, then we have to care about the entire lives of persons. A view according to which the scope of distribution concerns the entire lives of persons would have to be indifferent when presented with these two scenarios. The reason is that both scenarios contain an equal number of benefits for the person’s life as a whole. Parfit’s point so far is thus: There is scope for the application of distributive principles within people’s lives, not just across lives.

A slightly modified version of the example will illustrate this intra-personal distribution principle further. The two options we have to choose between could look like this:

(A2) giving Smith 10 benefits at the age of 40, or
(B2) giving Smith 4 benefits at the age of 40 and 4 at the age of 50.

Again, it is assumed the benefits are all the same.

Somebody who believes we should simply maximize benefits within our lives would have to favour A2 here. But somebody who values equal distribution within lives might opt for B2. Moreover, equality and maximization might not be the only competing principles about how to promote benefits within people’s lives. Especially with regard to some goods, professional accomplishment for example, many people might actually prefer a continuously ascending distribution or a bell curve instead of equality or maximization within their lives.

---

75 In Rawls, “The Independence of Moral Theory,” p. 18, Rawls describes his Kantian criterion of personal identity: “The ideal is that of persons who accept responsibility for their fundamental interests over the span of a life and who seek to satisfy them in ways that can be mutually acknowledged by others.” This seems to suggest that he might be in favour of applying distributive principles to whole lives. It is also possible to construct egalitarian views that take life stages into account, though. Daniels, “The Prudential Lifespan Account of Justice across Generations,” McKerlie, “Equality and Time” and “Dimensions of Equality”, Temkin, Inequality.
As Parfit insists,\textsuperscript{76} taking the temporal parts of people’s existence into account can sometimes yield the more intuitive answer to distribution problems. He illustrates this by appeal to a case like the following:

Two people are suffering a burden right now only one of whom we can help. The first person is undergoing a large burden right now. The second person suffers only a little right now, but suffered tremendous burdens in the past.

If we take their life totals as the basis of our decision we would have to help the person who is now suffering less. Many people will find this unintuitive. If we take their suffering at the present time as the basis, then we should help the person who is now suffering more. This would seem to be the better response to the case.

Parfit’s position does not necessarily imply that we should give exclusive weight to time-stage equality. Instead, one may also want to assign some relevance to both considerations and assign a certain amount of weight to equality between entire lives and equality at particular time-stages. This might be especially intuitive in cases other than emergency situations involving pain. If we have an educational toy to give away to one of two children in the developing world, we may well want to give serious consideration to how many such toys the eligible children have had a chance to play with in the past in addition to considering equality between them in the present.

The example also raises some further questions. Should we assign weight to equality between simultaneously lived life stages, such as Joe’s childhood and his mother’s adulthood, or should we also assign weight to equality between everybody’s childhoods, such as Joe’s

\textsuperscript{76} Parfit, \textit{Reasons and Persons}, p. 341.
childhood and his mother’s childhood? What about extending this thought to the childhoods of past and future persons? Not restricting such comparisons to the living seems to have the disadvantage that it can lower the prospects of the living or of future generations, if people were generally worse off in the past.

For now, only a couple of further clarifications should be added: First, Parfit’s claims do not extend “to society the principle of choice for one man,” as Rawls worries in the context of the passage quoted above. Quite to the contrary, he is not claiming that persons are maximizers within their own lives and that one should consequently choose maximization as the principle for entire societies. Instead, his view actually gives a lot of room to the thought that persons might actually prefer an appropriate distribution of benefits within their lives over maximization. So there is no argument from intra-personal principles to principles for societies.

Second, the Parfitian view that there is scope for distributive principles within people’s lives does not necessarily amount to the claim that people’s lives are not separate from each other. The view that people’s lives are not separate from each other is a view that has been attributed to utilitarians by Rawls, Gauthier and Nagel. They thought that utilitarianism leads to a conflation of the separateness of persons into one big mass person or super-person. This, however, would amount to the claim that there is no scope for distribution at all any more, because there would be only one unit left: the one mass person. If there is only one person, then distinctively egalitarian principles have no application at all. Parfit, however, believes that even the temporal parts of persons are to a significant extent separate from each other. This is the opposite of believing in conflation. For him there is not only

separateness between people, but also a certain degree of separateness between the temporal parts of these persons. He consequently believes that the scope for distributive principles increases, not that it shrinks to zero, because there is only one mass unit.

He even acknowledges that the increased separateness initially makes his view count against utilitarianism rather than in favour of it. The reason is that widening the scope for distribution has the potential of making distribution even more important. The way Parfit eventually wants to turn his reductionism into a view supportive of utilitarianism is through the further claim that reductionism supports giving less weight to distribution.

To figure out whether reductionism really reduces the weight of distributive principles, we have to consider their weight in both cross-personal and intra-personal cases. With regard to cross-personal cases, he gives one central reason why reductionism supports giving less weight to distributive principles: According to reductionism a further fact of personal identity does not matter. If that is so, then the question to whom something happens is not as pressing anymore and the question when something happens becomes more important. If identity matters, there are two facts that speak in favour of distributive principles: the fact of identity, and the relations of psychological continuity and connectedness between the various parts of the identical self. If identity does not matter, there is one fact less that speaks in favour of distributive principles.

With regard to intra-personal cases, Parfit presents an argument about the nature of compensation that initially actually increases the weight of distribution in these cases. (This is the first of two arguments about compensation he introduces.) The argument runs as follows.

Opponents of reductionism tend to believe someone who suffered a lot in the past can always be *compensated* for his or her past suffering by receiving a related benefit in the future. But if reductionism is true, we cannot say anymore that the person receiving the compensation will in the relevant way still be the same as the one that suffered the burden. To illustrate this, he discusses a scenario like the following:

A parent has to decide whether to impose a burden on a child.\(^8^2\) If the burden is imposed this will either lead to

(A) a benefit for the child’s own future self, or

(B) a benefit for another child’s future self.

The two children do not know each other.

Many people believe a benefit that the child will later enjoy herself will compensate her for the burden in the past. The same people might believe a later benefit for another child cannot compensate the burden of the first child. The reductionist view, on the other hand, would hold that the two scenarios have to be treated more alike. The reason is that reductionism treats the relationship of a child to her future self more like the relationship to another person.\(^8^3\) This means: Even if the benefit takes place in the first child’s own life there might not be full compensation. This decreased chance for compensation in turn implies that distribution should become more important in the intra-personal case.

This argument about compensation is characteristic for *Reasons and Persons*. It claims that there is *no guaranteed full compensation in intra-personal cases*. Such guaranteed full


\(^8^3\) Parfit, Reasons and Persons, p. 335: “A Reductionist is more likely to regard this child’s relation to his adult self as being like a relation to a different person.”
compensation would have required that identity matters.\textsuperscript{84} This argument is compatible with the view that there can be a reduced degree of compensation depending on the strength of our psychological relatedness across time. One may not expect full compensation though. We shall see further below that Parfit’s later use of the notion of compensation is different.

Disregarding the later shift in his conception for now, we can summarize that Parfit’s reductionism is supposed to support giving less weight than anti-reductionism to distributive principles in the cross-personal case, but so far turned out to give them more weight in the intra-personal case. Commentators have worried that this will support egalitarianism of temporal parts of lives rather than utilitarianism.\textsuperscript{85} In order to qualify this result we now need to take a closer look at Parfit’s reductionist position.

Under scrutiny, his view is ambiguous in at least the two ways discussed in the previous chapter. Before coming to final conclusions about the weight of distributive principles, we therefore have to take these ambiguities into account as well as how his later argument about quasi-compensation works. In the end, we will be able to assess how the various aspects of his reductionist position affect the weight of distributive principles.

\subsection*{3.2.1 Parfit’s Ambiguities}

As discussed in the previous chapter, one of the points Parfit remains ambiguous about is the relative importance of psychological continuity and psychological connectedness. He fluctuates between saying that connectedness matters the most and saying that continuity and connectedness matter equally. As pointed out before, on page 206 of \textit{Reasons and Persons} he writes:

\begin{verse}
\end{verse}

\textsuperscript{84} Parfit, \textit{Reasons and Persons}, p. 337.
Psychological connectedness is the holding of particular direct psychological connections. Psychological continuity is the holding of overlapping chains of strong connectedness. Of these two general relations, connectedness is more important in both theory and practice.

On page 301, though, his beliefs seem to have changed and he writes:

I believe that both relations matter. Others may believe that one matters more than the other. But I know of no argument for such a belief.

This is a surprising self-contradiction. Parfit is obviously ambiguous about the relative importance of the two notions. Our argument in the previous chapter suggested that it is not implausible to believe the notion of connectedness should matter more.

Parfit’s second ambiguity lies between the claim according to which our normative judgments will have to change due to the unimportance of a further metaphysical fact of identity and the view that they will have to change on the ground that continuity and connectedness matter to us in survival. Again, Parfit seems to have changed his opinion on the matter. In Reasons and Persons he is undecided. In “Comments” he seems to shift toward the view that the unimportance of identity is what has the decisive normative impact.

The choice between these latter two views—or between how much relative weight we want to give to each in a combined approach—seems to be relevant for normative matters. The reason is that each view would yield different results when considered on its own. To illustrate this, I will now take a look at the consequences of each view considered on its own.
If the lack of a metaphysical further fact is the more important consideration, then the units to which our moral principles should apply could become as short as momentary experiences. The impression that the various parts of my individual life are parts of my life would be based on a mistaken assumption about a metaphysical fact. According to reductionism there is no deep metaphysical concept that unites the various moments of my life into the life of a single person. If the relations of continuity and connectedness are the more decisive aspects, then the units to which our principles should apply remain temporally extended.

It was Daniels who first noticed that it is unintuitive to assume that the two positions will have the same normative impact when considered on their own. The problem he sees is the following: If Parfit’s claim that reductionism lowers the weight of distributive principles depends predominantly on the view that there is no metaphysical fact of personal identity, then the weight of distributive principles would have to be low even in cases in which there is no temporal disintegration of the self and we are dealing with strong continuity and connectedness instead. This can seem counter-intuitive. If there is strong psychological continuity and connectedness in a person’s life, why should he or she be treated as a mere succession of momentary mental incidents, and why should the weight of distributive principles between internally strongly related units be low?

To look at the problem in more detail, consider the following. On its own, denying the relevance of identity means that there is no relevant unity relation that turns people’s lives into a single whole. Instead we are dealing with a fundamental disintegration across time that applies in all cases. Now imagine a person who has a photographic memory of everything she

did as a child and who has remained true to all her intentions for the future. If the lack of the
metaphysical identity relation matters and we are always fundamentally disintegrated, then
this has to be true in all cases, including this one. Her adult self and her childhood self would
have to be treated as two disintegrated units. The lack of a metaphysical fact would determine
that the two stages are almost like two different people and that the most appropriate units for
distributive principles should be the various moments of the person’s life. There would have
to be scope for distributive principles within this strongly integrated unit and their weight
would have to be higher than on an anti-reductionist view. This seems unintuitive. Given the
strong internal relations this person does not appear very disintegrated after all.

Surely, we should say something else instead. We should say that in such cases the
strong presence of psychological relations matters a lot and should make it less important how
benefits are distributed between the various moments within the strongly internally related
unit. In terms of the intra-personal psychological relations this child and this adult are pretty
much like one person after all. Applying distributive principles to all momentary parts of a
person who remembers her entire past and has never abandoned any intentions for the future
seems to be rather unimportant. The weight of distributive principles applied to momentary
units, if they should be applied in this way at all, should be low in such cases. Of course they
could still be somewhat higher than on an anti-reductionist view that would not give any
scope to this kind of distribution at all. At the same time, their weight should be very
significantly lower than the weight that anti-reductionists would assign to distributive
principles between separate persons. So apart from the intra-personal cases where our
intuitions seem to depend on how integrated someone is, assigning exclusive scope to
distribution between momentary experiences would lower the weight of the principles
significantly in the cross-personal realm.
This does not convince us yet that distributive principles should not be applied to units of strong psychological continuity and connectedness and that the weight of the principles applied in this way should not continue to remain high, both in intra-personal and in cross-personal cases. Assigning predominant importance to psychological continuity and connectedness avoids the unintuitive result that our distributive principles should apply to all moments of all lives even in cases of persons with photographic memories. It would allow us to treat the person with the photographic memory as one distribution unit, which seems more intuitive. A very forgetful and cluttered person, on the other hand, could be treated as several units. The cluttered person might, for example, hardly remember any of the things he experienced and did as a young child. In his case, we could treat the child and the adult as two separate units. The lengths of the relevant distribution units would depend on the specific degree of intra-personal psychological relatedness that someone displays and not on the unimportance of a metaphysical concept, which applies across the board.

However, assigning predominant importance to psychological continuity and connectedness will also raise a problem. Parfit wants to be able to say that the weight of distributive principles declines overall. But why should the weight of distributive principles decline if the size of the units depends on strong psychological continuity and connectedness? Take the case of the cluttered person. His childhood self and his adult self are almost like two separate persons. There are almost no psychological relations. If anything, their faded relationship would seem to speak in favour of increased weight for distributive principles between intra-personal connectedness units.

The same would seem to be true for connectedness units of separate persons. If connectedness units contain the kinds of diachronic things in life that we find especially important at any given time within that unit, then the weight of distributive principles between
them should be particularly high in the cross-personal case as well. If the weight of
distributive principles can increase in these cases, then we may have found an additional
argument in favour of distribution sensitive theories rather than utilitarianism. Parfit’s view
would then speak in favour of time-stage egalitarianism, as his critics have worried.

These considerations show that Daniels’s objection is not complete. He makes it seem
as if the problem applies only to Parfit’s suggestion about the weight of principles. However,
it also applies to the suggestion about their scope. If Parfit’s claim about the scope of
distributive principles depends on the lack of a metaphysical further fact, then the scope
would have to be the same both for the person with the photographic memory and for the
cluttered person. This is counter-intuitive, though. In the case of the cluttered person, positing
intra-personal scope for distributive principles seems comparatively more plausible than in the
case of the person with the photographic memory.

Given these problems, one should want to pursue a different route. One might want to
go for a combined approach and actually give some weight to both considerations: the lack of
identity and the psychological relations. The lack of the further metaphysical fact of identity
would reduce the weight drastically in all cross-personal cases and would make it slightly
higher in intra-personal cases in comparison to a traditional anti-reductionist view. At the
same time, we could give some weight to psychological relations to accommodate the
conflicting intuitions about photographic and cluttered minds. Taken together, such a
combined reading might support utilitarianism on the ground that it reduces the weight of
distributive principles at least somewhat when considered on the whole. The main force in
Parfit’s favour would be the massive lowering of the weight through the lack of the further
fact in cross-personal cases.

In *Reasons and Persons*, Parfit is still hesitant about whether the lack of a further fact
should really change the units for distributive principles to all moments of all lives. He labels the position, and its consequences for the weight of distributive principles, an “extreme view.” In his later article “Comments” his position shifts though. In “Comments” he argues that the lack of a further fact should have this impact on the scope and that it will reduce the weight of distributive principles. In “Comments” he also adds, though, that temporally extended relations of care may have an additional impact. This might imply that he ended up endorsing the combined reading suggested here.

3.2.2 Agent-Neutral Quasi-Compensation: A Utilitarian Ideal?

Before coming to any final conclusions, we should still take a look at the new argument about compensation that Parfit introduces in “Comments.” This new argument continues to insist that compensation would require identity, but it also argues that there can be quasi-compensation whenever there is a care relationship between momentary selves—including the momentary selves of separate persons. By hypothesis we may assume that Parfit thought this later argument would be at least as consistent as his earlier view, would reduce the weight of distributive principles possibly even more, and would thereby lend more support to utilitarianism.

The argument in favour of his new combination of the “extreme view” with quasi-compensation works as follows: We imagine that a person branches like an amoeba. After the branching incident a time of scarcity begins. At least one offshoot is informed by an administrator that he will receive particularly few resources now. The reason: He has been compensated in advance by the original person’s life in luxury. The offshoot is outraged. He does not believe that he is compensated by the wealth of the original person. Psychological continuity and connectedness do not guarantee compensation, he insists.
Parfit agrees with the offshoot. In the example of a child’s burden in *Reasons and Persons*, Parfit still left open the possibility that there can be *some* compensation in accordance with one’s specific diachronic degree of psychological relatedness. Now he insists explicitly that not even strong psychological relations can do the trick.\(^{87}\)

The offshoot has two options for arguing against the administrator. First, he can argue that identity matters. He can then point out that he is not identical with the original person and should therefore receive as many resources as everyone else. This route is not open to Parfitians. They believe identity is not a deep fact and does not matter.

Second, he can argue that psychological continuity and connectedness cannot guarantee any compensation. Compensation would have required that identity play a role. Since he is only related to the original person through psychological continuity and connectedness, though, there is no compensation between him and the previous person. This is the answer that Parfit settles on.

Since the view that there cannot be any compensation across time at all is an “extreme view,” he introduces the notion of quasi-compensation at this point:

…because X loves his children, the benefits to them give him a kind of compensation.

Since it will not be he who receives these benefits, I shall call this “quasi-compensation.”\(^{88}\)

A little further down the page he specifies:

----

\(^{87}\) Wachsberg supports this position in “Identity, the Nature of Persons, and Ethical Theory.”

\(^{88}\) Parfit, “Comments,” p. 871.
Just as we can be quasi-compensated by benefits to those other people about whom we care, we can be quasi-compensated now by benefits to ourselves in those other parts of our lives about which we now care.\textsuperscript{89}

One might be tempted to think that the difference between his earlier view about compensation and this account of quasi-compensation is merely a matter of language, but this is not the case. The earlier account also left open the possibility that there might be a reduced degree of compensation depending on a person’s degree of psychological relatedness and the kind of special self-concern that goes with these relations. Quasi-compensation depends on care. Special self-concern and care are not the same concepts though.

First of all, care does not necessarily have to contain a self-biased element. We can care about things that have nothing to do with us. For example, we might care about the fact that there are starving babies on the other side of the planet. Second, Parfit’s example suggests the relevant relations of care can apply between separate persons such as a parent and a child. This is not a kind of relation that he takes to be relevant in *Reasons and Persons*. In *Reasons and Persons* he focuses on the kind of future-directed special self-concern that matters to us in survival, which includes such things as fear of oncoming pain,\textsuperscript{90} and on recollections of the past. Arguably, these kinds of relations cannot apply between a parent and a child—or only in a less immediate way. A father cannot literally feel his child’s pain, nor does he remember last night’s dinner table from the exact same spatial perspective. The notion of care that Parfit’s later account of quasi-compensation is based on, and the psychological relations that

\textsuperscript{89} Parfit, “Comments,” p. 871.
\textsuperscript{90} Parfit, *Reasons and Persons*, section 83.
the possibility for a reduced degree of compensation in his earlier account are based on, are substantially different.

I believe there can be a strong and a weak reading of this proposal in favour of quasi-compensation. The fact that it talks about parents and children makes it seem as if the passage endorses an agent-relative account of care.\textsuperscript{91} If this is true, the account would presumably not support utilitarianism, because utilitarianism aims to be agent-neutral. This would be a problem for Parfit’s agenda.

On the other hand, there could be a somewhat more agent-neutral reading of the passage that has not been pursued. If such a position works, we can use it as the basis of an agent-neutral understanding of quasi-compensation: the claim that one can generally be quasi-compensated by benefits to all others, irrespective of family or group membership. This could be a view that speaks in favour of utilitarianism.

In order to figure out whether such a view is possible, it makes sense to look at Parfit’s views about the notion of special self-concern in his earlier work once more. In \textit{Reasons and Persons}, Parfit argues that we have special self-concern for those parts of our future existence that we will be strongly psychologically related with. This is why we believe what matters is preserved in cases of branching. At the same time, Parfit believes our special concern for only loosely connected parts of our future can become almost as weak or as weak as our reason to be concerned about separate people.\textsuperscript{92}

This reasoning does not imply any agent-relative element for concern or care about others. In fact, it implies an \textit{agent-neutral} account of care about separate others. Our reason to

\textsuperscript{91} Jeske argues in “Persons, Compensation, and Utilitarianism” that this is the view Parfitian reductionism supports. Parfit’s one sentence remark on p. 337 of \textit{Reasons and Persons} supports this position. However, there is no similar exclusion of the agent-neutral reading in his more explicit treatment of the matter in “Comments.”

\textsuperscript{92} Parfit, \textit{Reasons and Persons}, p. 335, 337.
be concerned about our own loosely related future selves and our reason to be concerned about all separate others is about equal.

If we have equal reason to care or to be concerned about all others and can even be quasi-compensated by benefits to those others, the prospects of utilitarianism’s agent-neutral and comparatively distribution insensitive maximization principle seem to look rather good. Combining Parfit’s views on what we have reason to be concerned about from *Reasons and Persons* with his later view about quasi-compensation would therefore to do work in favour of utilitarianism.

One may object at this point that Parfit’s later view about quasi-compensation, especially in an agent-neutral form, is counter-intuitive or even unjust. Many people will deny that they can be quasi-compensated for their burdens if total strangers receive resulting benefits. We usually accept the idea that parents can be quasi-compensated by benefits to their children. But should we believe the agent-neutral thesis that benefits to total strangers will accomplish the same? Critics will insist that agent-neutral quasi-compensation is an *ideal* rather than a fact. They could further insist that it is this ideal, rather than the reductionist view about persons, that is doing the work in favour of utilitarianism. Furthermore, the ideal as such might presuppose most of what is needed in order to believe in utilitarianism.

The charge that the ideal, rather than the reductionist view about persons, would be doing the work in favour of utilitarianism can be further supported as follows: Reductionism actually speaks against the belief that we could receive real compensation by benefits to our own future self. Compensation would have required identity, but identity does *not* matter. Reductionism consequently says that our reason to be especially concerned about our own future self gets reduced. We only have a strong reason for special self-interested concern within the time period within which we are strongly psychologically related. Outside of this
time frame, our reason for self-concern is reduced and depends on our specific degree of psychological self-relatedness. This view only supports a form of egoism with a discount rate on the basis of one’s degree of psychological self-relatedness or egoism of life stages, rather than anything else. Only if we add the ideal demand that we ought to experience care and quasi-compensation in regard to separate persons (as well as to only remotely related parts of our own future) will the view turn out to support utilitarianism.

However, why should an appeal to ideals not be open to us here? Maybe it is acceptable if utilitarians presuppose an ideal criterion of selfhood and compensation. Their ideal would envision selves that achieve quasi-compensation through benefits received by others that promote overall utility, even if they themselves are facing a distributive burden as a result of this. If the occurrence of such quasi-compensation in people is not actually the case right now, then maybe we ought to adjust our dispositions accordingly and thereby change the facts.

Competing views have also endorsed ideal conceptions of the self. In a context that I will discuss in more detail in the last chapter, Rawls endorses a “Kantian criterion of personal identity” that is an ideal people should strive toward rather than a fact. He writes:

The ideal is that of persons who accept responsibility for their fundamental interests over the span of a life and who seek to satisfy them in ways that can be mutually acknowledged by others.93

Utilitarians could pursue a similar route and use an ideal conception of the self in support of their own view. Strategies that are open to others should be open to them.

On the other hand, this strategy might involve giving up one of the strongest aspects of Parfit’s project, namely, the claim that we can proceed from comparatively non-normative facts about persons to a particular moral theory. As Stephen Darwall points out in “Scheffler on Morality and Ideals of the Person,” it is not a natural feature of the utilitarian position to endorse an ideal of the person. In fact, the compelling aspect of the utilitarian position might be precisely that it does not endorse an ideal of the self and that it seems to move from non-normative facts about persons to a moral principle. A reading of Parfit’s view that assumes an ideal of the self would lose this advantage, which probably motivated his entire investigation into the nature of persons.

Moreover, simply demanding that we ought to adjust our dispositions so that we tend to experience care and quasi-compensation with regard to all other people might not actually be an independent argument in favour of utilitarianism. In fact, it might merely be a different way of demanding that we should adjust ourselves so that we will have utilitarian dispositions. Where would the extra argument for utilitarianism lie in this case? The demand to adjust our dispositions in order to move closer to a utilitarian ideal of the self might be just a different way to express a simple demand to become a utilitarian.

3.2.3 A Holistic Reading of the Early View

I am going to suggest that a holistic reading of Parfit’s earlier view in Reasons and Persons will fare better and can be seen as yielding an optimal version of a reductionist utilitarian position. We have to turn to his earlier view in order to rescue the support for agent-neutrality without introducing a utilitarian ideal of the self.

---

In *Reasons and Persons*, Parfit entertains both the view that the lack of a further fact of identity has implications for distributive principles and the view that diachronic psychological relations have implications for distributive principles. He is also ambiguous between the views that the relations of continuity and connectedness matter equally and the view that connectedness matters more. To flesh out the implications of a view that combines all these elements, I am going to look at the implications for the weight of distributive principles that each of these features will have on their own.

The view according to which the lack of a further fact of identity has implications for distributive principles increases the scope and lowers the weight of distributive principles, but it does so at the expense of counter-intuitive consequences for persons with photographic memories. I have mentioned earlier that it seems possible to out-balance this problematic consequence by endorsing a combined account according to which diachronic psychological relations have implications for distributive principles in addition to the lack of a further fact of identity. More specifically, we could say that the lack of a further fact decreases the weight of distributive principles quite drastically across the board. At the same time, we could hold that the diachronic psychological relations give some weight again to the distribution between strongly psychologically related units. This will prevent the weight from declining overly drastically in all cases and will accommodate intuitions about cluttered and photographic minds.

The reason the view according to which the lack of a further fact of identity has implications for distributive principles would reduce the weight quite drastically is that a proper distribution between all moments of all lives is rather unimportant to us—at least when dealing with ordinary persons rather than extremely cluttered ones. It would factor in all sorts of irrelevant timing issues: If a particular benefit arrives a few moments later than scheduled
this could always be seen as an injustice. Clearly, we don’t find this form of injustice generally bothersome at all.

Nevertheless, the weight of this kind of injustice might not decline all the way to zero. Occasionally, exact timing might matter after all. Also, as long as there are any units that we can distinguish, even just momentary experiences, it might still seem somewhat bothersome if some of them are full of bliss and others are filled with pain. This means the urge for distributive equality could still arise. We might not want some momentary experiences to be extraordinarily pleasant while others are only full of misery. Overall, this view on its own would therefore reduce the weight of distributive principles very drastically, but not quite all the way to zero.

Now we still have to answer the question of whether the diachronic relations might not increase the weight enough again in order to entirely outbalance the effect of the lacking further fact of identity. To figure this out, it is important to distinguish between the individual forces of psychological connectedness and the force of psychological continuity on its own, before attempting to construct a combined view.

What would the picture look like if connectedness mattered on its own? In contrast to the implications of the lack of a further fact, we would then be dealing with units that are temporally extended. What would matter would be those units that are characterized by direct diachronic psychological relations such as memories or intentions. Arguably, the time periods over which people are psychologically connected can be fairly long. Older people usually still remember a lot of what they did when they were teenagers. Likewise, teenagers are starting to form intentions that can shape their lives for decades to come. But at the same time, these connections never hold completely. Overall, connectedness is a matter of degree. By the time people are eighty, they usually have irrevocably forgotten most of the lunches they had as
young adults. The length of the units we are willing to view as strongly internally connected will depend on further details of the view.

In any case, we will be dealing with units that are longer than just a moment. This means that we would not have to factor in all sorts of irrelevant timing issues, as the previous view would. This seems to bear the potential for making distributive issues more important than they would be if the lack of a further fact were all that mattered. At the same time, such a view could factor in all sorts of issues that a momentariness view could not account for and that we tend to find important, for example the sense of accomplishment after the completion of a long term project or the merits of a long term relationship.

It might be particularly bad if there is inequality between such units of strong connectedness. The reason is that they contain the kinds of things in life that persons tend to be especially concerned about and that they value the most in their lives. In contrast to a continuity unit, within one such connectedness unit persons are still especially aware of and related to the things they find important at all the other times within the unit. A connectedness view is therefore bound to find a proper distribution very important. Depending on how important it makes distribution it might even lead to egalitarianism of life stages, which would then seriously threaten Parfit’s agenda of finding support for utilitarianism instead.

However, we may still try to find other arguments that might offset this conclusion and establish that the weight of the principles will decline after all. One such argument could be the following. If strongly connected units matter the most, then we can say that our stream of consciousness will actually “live through” a number of such units. We do not really cease to exist once we have forgotten a lot about our childhood. Our stream of consciousness continues. We are continuously sliding into the next unit of strong connectedness. If we are doing worse than our fellow citizens right now, we may find comfort in the thought that our
next unit of strong connectedness might actually do better. There is hope that within the “life after” the current connectedness unit, so to speak, things will improve. Moreover, we will have to realize that at some point in the future we may have forgotten that we were once so bothered about being worse off than others right now. A future connectedness unit of ours might be blissful and entirely untouched by our current situation. It might not even remember the present inequality anymore.

An important feature of this argument is that it assumes continuity matters in addition to connectedness. Otherwise, we would have no special ground for believing that our stream of consciousness will live through further connectedness units. Those further connectedness units would have to seem simply and entirely unrelated to us. The work of this argument is done by the implicit assumption that continuity matters as well. This is an interesting result. Constructing a reading of the connectedness view that moderately reduces the weight of distributive principles, even if we believe that perceived inequalities are worse, seems to depend on the idea that continuity matters also.

This leads us to ask what impact the view that continuity matters would have on the principles when considered on its own. It is actually possible for people to remain psychologically continuous over the course of their entire biological lives. As a consequence continuity seems to come rather close to a more traditional whole lives view in terms of the impact on distributive principles. Given this comparatively neutral impact, continuity as such can be seen as outbalancing some of the weight increase that the connectedness view could lead to on its own. This means that the increased weighting of connectedness will probably not outweigh the weight loss of the view that the lack of a further fact produces.

In combination, the various elements of Parfit’s reductionism in *Reasons and Persons* can therefore be seen to support a moderate weight reduction for distributive principles—
without running into the counterintuitive consequences the “extreme view” runs into when considered on its own. This could give some relative support to utilitarianism in comparison to competing theories that give more or even absolute weight to distribution considerations. This is an encouraging result for Parfitian utilitarians.

However, before concluding we still have to consider one final objection against the combined reading of Parfit’s earlier reductionism in *Reasons and Persons*. We may still ask whether the weight reduction that this combined view supports will really lead to utilitarianism rather than a form of discount rate egoism or egoism of life stages. A loss of weight for distributive principles might not necessarily lead to utilitarianism; it might also lead to egoism. More specifically, it might lead to a view according to which not just distributive principles, but just about anything outside one’s own current relatedness unit loses its importance. How does the combined view support the *agent-neutral* utilitarian maximization principle, rather than the view that apart from the welfare of my current relatedness unit nothing is important?

I suggest that both an intra-personal as well as a cross-personal understanding of the transitive notion of continuity is the best answer available. To get from a form of egoism to utilitarianism, we need an agent-neutral element that shows the similarity between our relationship with other parts of ourselves and our relationship with separate persons. In his later work, Parfit tried to provide the required element by introducing the notion of quasi-compensation. Quasi-compensation can occur both intra-personally as well as cross-personally. It gives people a reason to care about others that is similar to the reason they care about themselves. However, the problem of this solution is that critics will insist that the underlying notion of care is not agent-neutral.
Continuity suggests that I can be indirectly related to an entity in the distant future, even though I cannot imagine yet what that being’s mindset will be like. Even though I might now already have certain intentions for the future, I already know that my very distant future self might not hold these intentions anymore. Nonetheless, I might be at least somewhat egoistically concerned about this being.

As the passages quoted above seem to suggest, Parfit believes the self-concern that comes with this indirect relationship can become as low or almost as low as my reason to care about any separate person. If this is true, then we have a way of emphasizing the idea that my relationship to my own remote future self is essentially like my relationship to separate others, including distant strangers. The kind of indirect relation that I have toward my own distant future self might be in a sense just like the relations that I can in principle share with all separate others. This could get fleshed out in the following way. I do not seem to have access to the mindset of my distant future self right now and cannot even anticipate that mindset to a significant extent. Nonetheless, I am sure it will possess consciousness, will have fears, beliefs, desires and intentions, and I genuinely hope it will fare well. Identity gives me no special reason to assume these things in my own case, if a further fact of identity does not exist and if identity does not matter. Since I nonetheless seem to believe all these things about my distant future self, there seems to be no reason why I should not likewise assume them in the cases of all separate others. In both cases there is a peculiar lack of epistemic access to another mind and I am nonetheless entirely sure these minds are and will be there. Since I am nonetheless concerned about this mind in my own future case, it would seem consistent to be concerned about all other minds as well. The transitive concept of continuity as an anticipation that there are minds that I cannot access right now, and the fact that I nonetheless
genuinely hope they will fare well in the case of my own future selves, might be able to support an agent-neutral reason for care about all others.

If this does not seem compelling there also seems to be a second route of using the notion of continuity to rescue agent-neutrality. There might be some more substantial psychological relations that can shared by separate persons. For example, beliefs, desires and intentions can in principle be shared and mutually caused by separate persons. If it is at all plausible to assume that these relations can form transitive continuous chains, then we might have likewise found an element that links self-concern with other-regarding concern and that might be at least somewhat agent-neutral. A more detailed discussion of this option will be part of the following chapter.

Initially, there are at least two problems with this proposal. First, it assumes that there could be continuity among all persons. It also assumes that this continuity is an all-or-nothing relation. While this is how it has traditionally been construed, we may not simply assume this traditional habit is immune to criticism. Parfit himself suggests to speak of continuity “if the number of direct connections, over any day, is at least half the number that hold, over every day, in the lives of nearly every actual person.” This shows that the use of continuity as an all-or-nothing relation might be a rather arbitrary way of drawing a line in a spectrum of degrees. Jeff McMahan, for example, draws the consequence and suggests construing

95 Parfit, *Reasons and Persons*, p. 206. To give some more context, here is a longer version of quote: “Psychological continuity is the holding of overlapping chains of strong connectedness. (...) Since connectedness is a matter of degree we cannot plausibly define precisely what counts as enough. But we can claim that there is enough connectedness if the number of direct connections, over any day, is at least half the number that hold, over every day, in the leaves of nearly every actual person. When there are enough direct connections, there is what I call strong connectedness.”
continuity as a matter of degree. If continuity can admit of degrees, then this view might not turn out to be agent-neutral.

Second, the suggested reading of *Reasons and Persons* assumes that the relevant intra-personal and cross-personal continuities that matter are of the same character. This may be disputed, however. It might well be the case that there are particular kinds of intra-personal psychological relations that never apply in cross-personal cases. For example, it is often said that we have a peculiar and very distinctive first-person access to our own mental states that we do not have to the mental states of anyone else. If this matters to a significant extent, then the needed link or symmetry between intra-personal and cross-personal cases is at danger of getting relativized. There would then be a particular kind of psychological relationship that never applies in cross-personal cases. This could lend support to egoism once more rather than to an agent-neutral concern for all others.

Finally, we may ask whether Parfit’s plan of defending utilitarianism against the attacks by Rawls and others has been successful. According to these attacks, utilitarianism disregards the separateness of persons and conflates all persons into one. Parfit argued that reductionism supports a bigger scope for distributive principles and reduces the weight of distributive principles. Does this response work?

With regard to the bigger scope for distributive principles, there seems to be a way in which Parfit’s project was successful. He has shown that a reductionist utilitarian does not conflate the separateness between persons, as the famous charge by writers such as Rawls, Gauthier and Nagel suggests. Instead of creating a moral unit that is bigger than a single

__________________________

person, such as a “super-person” or a “mass person,” Parfit created moral units that are *smaller* than a person, namely person-stages or momentary selves.

However, many people will take this to be a merely defensive point and not a positive argument in favour of utilitarianism. Parfit’s real argument in favour of utilitarianism was going to be that distributive concerns should lose weight. Utilitarianism is a view that is supposed to give no intrinsic weight to distribution at all. Critics will therefore insist that a complete defence of utilitarianism would require showing that distributive concerns have no weight at all. Unless Parfit managed to show that distribution should have no weight at all, it is open to critics to insist that his defence has not worked out.

I have argued that even on the most extreme reductionist aspect—the cross-personal momentariness view—the weight of distributive principles would never decline to zero. Even an extreme reductionist would have to maximize the happiness of *something*, for example of momentary experiences. This will always leave room for an incentive to give at least a small amount of weight to distributive principles. This means that the most demanding critics of utilitarianism will not be satisfied.

There is only one condition under which the weight of distributive principles would really decline to zero for a reductionist. It is the condition that Parfit argues against: the view that there is only one big mass person around. In this case, the scope for distributive principles would shrink to zero, because there would be only one unit left for which we could distribute. In this case, we would have to give zero weight to distributive principles, because it is clearly unimportant to try to distribute things among one entity. But this is not a view Parfit endorses.
Parfit endorses a view that will always result in at least a small amount of weight for distributive principles. The reason is that he believes in a significant degree of separateness within the lives of identical persons, not just between lives. In other words, he does not believe that there is a big mass self. And as long as there is any separateness, there is an incentive to distribute benefits fairly. Since utilitarianism is not supposed to give any intrinsic weight to distribution, this makes it seem as if Parfit’s real accomplishment was not a defence of utilitarianism as he initially thought. He ended up defending another, maybe better form of consequentialism that gives at least a small amount of weight to distributive concerns.

While Parfitians will always have to give at least a small amount of weight to distributive principles, there is considerable wiggle room with regard to how much weight they will give. The answer to this question will depend on a number of more detailed issues. They have to decide how important they find the lack of an irreducible metaphysical fact of personal identity as well as how important they find introspective psychological continuity and connectedness and possibly certain cross-personal relations. Depending on the answers to these questions, the relevant units for distribution will vary. For example, if all that matters is the lack of the further fact in all cases, including cross-personal ones, and we would not take any diachronic connections into account, then the resulting form of moral consequentialism would give distributive principles relatively little weight. On the other hand, if we believe that psychological continuity and connectedness matter as well, then the resulting form of consequentialism would be free to give distributive principles somewhat more weight.

However, there remains a possibility of rescuing the idea that Parfit successfully defended utilitarianism rather than some another form of consequentialism. The
presumption that the defence of utilitarianism would only be successful if the weight of
distribution would decline to zero could be wrong. The upshot of the reductionist
enterprise might not have been that utilitarianism is justified in giving no weight to
distributive principles. Rather, it might have been to show that utilitarianism is justified in
giving them less weight than, for example, Rawlsian egalitarianism.

Presumably, this is all that Parfit needed or wanted to accomplish. Under the
conditions that some form of reductionism is true and that it is not irrational to believe
reductionism will result in less weight for distributive principles, Parfit’s initial project
can be seen as relatively successful. In comparison with Rawlsian egalitarianism the
picture would subsequently look like the following: Reductionist consequentialists who
believe the most important aspects of personhood to be such relations as psychological
continuity and connectedness could give almost as much weight to distributive principles
as a Rawlsian egalitarian. The point that keeps them from giving as much weight as a
Rawlsian would be the belief that the lack of a metaphysical fact has some force as well.
Reductionist consequentialists who believe the lack of a metaphysical further fact matters
the most and who do not assign very much importance to diachronic psychological
relations would have to give much less weight to distributive issues than a Rawlsian
egalitarian.

This train of thought can be further supported by the following consideration. It
consists in arguing that giving a small amount of weight to distributive principles is
compatible with giving them no intrinsic value. Presumably even utilitarians would have
to create equality if this would promote the general happiness. But of course this does not
mean that they assign intrinsic value to equality. It merely means that they give
instrumental value to equality—for them, it is a tool for creating more general happiness.
The term “weight” as Parfit uses it throughout his treatments of this topic does not entail
the view that the value of equality is independent of the potential to create happiness. For
utilitarians equality can have value if and only if it promotes overall happiness. If equality
relatively regularly leads to more happiness, then at least rule-utilitarians would be willing
to assign a modest amount of instrumental weight to equality across the board.

Finally, the discussion has also yielded another point that has the potential to serve
as a response to people like Gauthier, Nagel and Rawls: the fact that people can
reasonably take units that are shorter than a whole life as the relevant units for distributive
concerns. Taking separate lives as the relevant units is not as neutral as Rawls and others
might have assumed.

Also, Rawls thought that the restriction against balancing should only apply to
separate persons. Subsequently, he never put any restrictions on *intra-personal balancing
acts*. For him, persons are always free to maximize within their lives. There are no self-
regarding moral obligations that prevent this between the temporal parts of the self.
Instead, self-regarding choices are matters of preference that are not of moral character.
For Parfitians, on the other hand, such self-regarding choices can enter the realm of
morality. For them, it can be an *immoral* violation against one’s own self to severely
exploit some stages of one’s life in favour of others. There is such a thing as treating
*oneself* unfairly.
4 Egoism, Morality and the Separateness of Persons
It is a widespread belief that rational self-interest or egoism is fundamentally at odds with the demands of morality. Even though egoism and morality might sometimes overlap, many believe that this will only be due to the egoists’ instrumental urge to make a good impression on others—which will in the long run benefit themselves.

In this chapter I am looking at two arguments from personal identity that are supposed to reduce the tension between self-interest and morality. The earlier argument by Derek Parfit concludes that the scope within which strong self-bias is plausible is much smaller than we tend to think. The more recent neo-Aristotelian proposal by David Brink claims the scope is a lot bigger than we tend to think. As I will lay out, both views hope to challenge the conviction that self-interest and morality are fundamentally at odds in the way in which Henry Sidgwick described it. Parfit’s account does so by saying that our diachronic psychological connectedness is generally rather low and that strong self-bias is only plausible in cases in which it is high. Brink hopes to accomplish this goal through his “metaphysical egoism” according to which friends and even more distant others stand in similar relations to us as our own future selves and we consequently have non-instrumental reason to be concerned about them.

In this paper I discuss the earlier argument by Parfit and provide an analysis of the central features and problems of Brink’s account. Finally, I introduce a hybrid approach that neither of them has anticipated. The hybrid approach seems to do more work in terms of easing the tension between egoism and morality than the suggestions by Parfit and Brink.

4.1 Sidgwick: The Separateness of Persons makes Self-Bias Rational

The background of both suggested arguments is the position that it is the separateness of persons that makes egoism rational. This separateness is supposed to be a precondition of the
belief that it is always better if I myself receive a good than if someone else does. When the good ends up in my life there is allegedly something more immediate or direct taking place from my point of view that is unrivalled by cases in which others receive a similar or even a larger good.

Sidgwick formulates the idea as follows:

It would be contrary to common sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently “I” am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals: and this being so, I do not see how it can be proved that this distinction is not to be taken as fundamental in determining the ultimate end of rational action for an individual.\textsuperscript{97}

According to this passage, common sense holds that the fundamental separateness between persons makes it rational to be more concerned about oneself than about anyone else. This is an argument from the nature of the self to the plausibility of rational egoism. Nonetheless, Sidgwick also believes that egoists can sometimes be converted. If they can be brought to accept that their own good is merely a part of the universal good, then it also becomes possible to convince them that from the viewpoint of this universal good their part is no more and no less important than any other part. They would subsequently have to care about the good of other persons in addition to their own. However, if the egoists deny this and insist that

\textsuperscript{97} Sidgwick, \textit{The Methods of Ethics}, p. 498.
their own good is all important while the good of another is never in the same way good for them, then there will be no way of proving them wrong, Sidgwick believes.\textsuperscript{98} As a consequence the dualism of practical reason arises: While it is rational to view one’s own good as a mere part of the universal good and to act in accordance with morality, it also remains rational to act in accordance with the self-interest theory.

\section*{4.2 Parfit: The Right View about Survival Undermines Rational Egoism}

Both of the arguments that I will now take a closer look at claim that the dualism of practical reason is less fundamental than one might think. According to Parfit identity does not matter. What matters instead are psychological continuity and connectedness with any cause. This view will undermine egoistic self-interest, he believes. More specifically, he claims that it undermines a central requirement of the egoist’s self-interested position—the requirement to be equally concerned about all parts of one’s life.\textsuperscript{99} Sidgwick formulated this feature of egoism as follows:

\begin{quote}
…my feelings a year hence should be just as important to me as my feelings next minute, if only I could make an equally sure forecast of them. Indeed this equal and impartial concern for all parts of one’s conscious life is perhaps the most prominent element in the common notion of the rational…\textsuperscript{100}
\end{quote}

\textsuperscript{98} Sidgwick, \textit{The Methods of Ethics}, pp. 420-21.
\textsuperscript{99} Parfit, \textit{Reasons and Persons}, p. 313.
\textsuperscript{100} Sidgwick, \textit{The Methods of Ethics}, p. 124
The right view about what matters, so Parfit’s argument goes, makes it implausible to have this equal concern for all temporal parts of one’s existence. Instead, it would make it plausible to have a discount rate for self-interest that depends on the degree of psychological relatedness between the temporal parts of the self. It would not be irrational to be less concerned about parts of one’s self that one is not very closely related to right now, such as parts that lie far in the future.

For example, thirty years in the future one’s psychology will have changed dramatically and one’s recollections of the present will have diminished considerably. In fact, most people will have forgotten most of their daily affairs. Consequently it is supposed to make somewhat less sense to have special self-interested concern for the being thirty years from now than for the being today or tomorrow.

If this is our attitude, though, then we do not care equally about all parts of our conscious life anymore and one of the most important aspects of the rational egoist’s position has fallen apart. The right view about what matters subsequently speaks against rational egoism. According to Sidgwick, egoism requires an equal amount of concern for all parts of one’s self. But the view that receives support from Parfit’s view about what matters is not like this. It supports a discount rate for special self-concern that depends on our degree of psychological relatedness with the other parts of our self.

4.2.1 Consequence: A Smaller Scope and Less Weight for Self-Bias

If all plausibility of self-bias is gone, there cannot be any tension between self-interest and morality. In cases in which self-bias is plausible to a very low degree, there can only be a very low tension between self-interest and morality. Let’s see how this plays out in detail.

There are at least three views against time-neutral self-interest that Parfit considers: the view that both psychological continuity and connectedness matter, the view that
connectedness matters more, and the view that identity does not matter. Psychological continuity is thought of as a chain of overlapping direct connections, such as recollections of the past. The exact implications of the discount rate for the tension between egoism and morality depend on how much relative weight one assigns to each of these ways of fleshing out the reductionist position.

If one gives primary importance to the idea that identity does not matter, there will never be a special reason to have any diachronic self-interested concern for any future part of one’s own life. The fact that a particular being in the future will be me rather than someone else will never give me a special reason to care. Subsequently, a tension between morality and self-interest could not really arise, since I never have special reason to care more about myself in the first place.

If one believes that connectedness is the most important relation for practical questions, the picture turns out different. Arguably, there are sometimes no direct psychological connections at all left between an eighty year old and his or her childhood self. The senior might not remember anything at all about his or her life at the age of four and the four year old presumably never had any specific plans about what to do at the age of eighty. The discount rate for special self-interest could therefore decline all the way to zero. Whenever the discount rate for self-interest declines to zero, there cannot be any tension between self-interest and morality at all anymore. But of course in many cases the ground for

For his deliberations about the view that the lack of identity matters for self-interest, see Parfit, Reasons and Persons, section 102. For considerations in favour of the view that connectedness matters more than continuity, see Parfit, Reasons and Persons, p. 206. For thoughts in favour of the view that connectedness and continuity matter roughly equally, see Parfit, Reasons and Persons, p. 301.

Momentary selves reductionism is the view that Parfit shifts toward after Reasons and Persons where he argues for the discount rate. This might be the reason he never explicitly points out that we will not have a discount rate if the view that we are momentary selves is what matters or what matters the most for normative judgments.
self-bias will still be present to a significant extent. For example, the connectedness between a 40 and a 50 year old will be somewhat reduced, but not all the way down to zero. In such cases self-bias is still plausible, albeit to a reduced extent. Subsequently a tension between self-bias and morality can arise, but only with reduced strength.

If one believes that both continuity and connectedness matter, then things will again look somewhat different. Even if the eighty year old has forgotten all about his or her four year old, there will still be a continuous chain of memory connections binding them together. This means the reason for special self-interest would persist on the basis of continuity. However, the degree of plausible self-bias would still decline according to the connectedness discount rate. The reason is that the view says the conjunction of continuity and connectedness matters. Since connectedness is a matter of degree, the conjunction will have to be a matter of degree. If only continuity mattered and under the common assumption that continuity is an all-or-nothing-relation, the discount rate would disappear. But this is not a view Parfit suggests. Parfit’s texts leave room for the interpretation that both continuity and connectedness matter, but not for the view that only continuity matters. On the conjunction view, there remains a tension between self-interest and morality. But in most cases, the plausibility of self-bias will be reduced according to the discount rate. It would usually be less strong than Sidgwick supposes when stipulating that egoists have reason to be equally concerned about all parts of their existence.

Brink has objected that Parfit’s view does not in fact establish that there will under all circumstances be a discount rate. The reason is allegedly that the presence of continuity establishes that all parts of people’s selves are equally parts of their existence as a whole. If they are all equally parts of their existence, it would have to be rational to be equally concerned about them. Subsequently, the tension between egoism and morality would be
equally strong in all cases and would not be reduced overall. However, this argument
overlooks two facts. First, it overlooks what Parfit really believes. He does not simply believe
that continuity matters, he believes the conjunction of continuity and connectedness matters.
Since connectedness is a matter of degree, so is the conjunction. Someone holding this view
would therefore not say that all temporal parts of the self are equally parts of the self. Second,
the view that only continuity matters and that therefore all parts of one’s self are equally parts
of one’s self is actually held by Parfit’s less reductionist opponents. For example, Sydney
Shoemaker believes that continuity matters and that there is an overarching explanatory
substance or identity relation that remains the same across time as long as there is such
continuity. Parfit acknowledges that this might be true, but he does not believe that these
things matter for practical purposes.

On two of the three possible readings of Parfit’s view there will be a discount rate for
self-interested concern. If any one of these readings is true, then the scope and the weight of
self-bias are smaller than Sidgwick thought. The view that identity does not matter does not
give rise to a discount rate, but neither does it provide a reason for diachronic self-bias. Both
the view that connectedness matters, and the view that continuity and connectedness matter,
will feature a discount rate and will subsequently reduce the tension between egoism and
morality. The view that only continuity matters will not feature a discount rate, but is not a
view that Parfit suggests.

One may nonetheless ask how the picture would turn out if continuity matters the
most. If continuity matters the most, we first have to ask whether continuity is to be seen as an
all-or-nothing relation, as it is generally construed, or as a matter of degree, as some authors
have suggested.\textsuperscript{103} If it is a matter of degree, then there will still be a discount rate. If
continuity is an all-or-nothing relation, then the reductionist picture can in fact turn out in the
way Brink’s criticism suggests: It would always be rational to be equally concerned about all
parts of one’s own life, just as Sidgwick described the egoist’s position.

One more clarifying remark about discount rates is in order. The discount rate
discussed depends on one’s degree of diachronic psychological relations. Indirectly, a
discount rate will also often result in a discount rate about time. The reason is that the
psychological connections will usually be lower if the two psychological states in question are
temporally far apart. Likewise, the relations will usually be higher if two psychological states
are closer in time. But the indirect temporal discount rate is not a necessary consequence of
this view. Brain damage, diseases or other forces may well decrease the psychological
relatedness between psychological states that are close in time. So primarily, this is a discount
rate with regard to the degree of psychological connectedness or continuity in conjunction
with connectedness.

We may still ask which one of two features actually causes the tension between self-
bias and morality: the special relatedness with one’s own self, or rather the lack of such
special relatedness with others. If it is the special relatedness with one’s own self, then Parfit’s
suggested reductionist position will in fact reduce the tension between egoism and morality.
However, if the source of the tension is the lack of special relations to others, then there will
now be a tension between disinterest and morality. Just as there would be no reason to care
about one’s distant future selves, there would also be no reason to care about others. This
negative aspect of egoism remains unaltered by Parfit’s theories as described so far.

\textsuperscript{103} For example, McMahan suggests treating continuity as a matter of degree. McMahan, \textit{The Ethics of Killing}, pp. 48-55.
As a response to this, we have to entertain the following considerations. So far, we have only addressed what Parfitian reductionism implies for the relationships among the various temporal parts of one’s own self. But one may also wonder what reductionism would imply for the relationships between friends, members of the same town or total strangers. Does he assume that the psychological connectedness between separate persons is always low or even zero? It seems as if he remains ambiguous about the answer. If such cross-personal psychological relations matter, then even those who are self-biased on the basis of psychological continuity and connectedness would sometimes have reason to care about others. In fact, there might now be a discount rate for other-regarding interest in addition to a discount rate for self-regarding interest. On page 319 of *Reasons and Persons* Parfit very briefly hints at this possibility. In “Comments” he makes a related point in terms of compensation. As discussed in the chapter on consequentialism, he argues that a father can be “quasi-compensated” by benefits for his son. I have argued that quasi-compensation is unlikely to turn out agent-neutral, which will endanger Parfit’s utilitarian agenda. People tend to be more concerned about their close ones than about strangers. Instead of endorsing an account of quasi-compensation I have suggested to treat the notions of connectedness and continuity as both intra-personal and cross-personal notions. A cross-personal notion of continuity might provide the agent-neutral reason to care about others that utilitarianism needs.

This proposal leads us to the second argument from personal identity that is supposed to bridge the gap between egoism and morality. According to this argument the proper way of thinking about personal identity supports rather than undermines rational egoism, but reduces

---

the gap between egoism and morality nonetheless. This argument is based on the idea just raised: that we can stand in special psychological relations to other persons and that these relations matter. According to Brink they give us a self-interested and non-instrumental moral reason to care about others as our extended selves.

4.3 Brink: The Right View about Persons Supports “Self-extended Egoism”

In ordinary life, people sometimes say that certain important others turned them into what they are. According to this way of speaking we are not entirely self-directed isolated individuals, but rather parts of a more extended psychological network. Brink’s view takes this understanding of personal identity literally. It emphasizes psychological relations that do not only apply between the various temporal parts of one life, but also between what we ordinarily think of as separate persons. According to his account, we are fundamentally intertwined with others. The reason is that the psychological relations that are relevant for the personal identity of individuals can in principle also hold between significant others or “extended selves,” as he puts it. If our relationship with others is consequently just like our relationship with ourselves, then the tension between morality and self-interest will cease to be fundamental.

Brink’s account of a person is:

On this view, a particular person consists of a series of psychologically continuous person stages. A series of person stages is psychologically continuous just in case contiguous members in this series are psychologically well connected. And a pair of person stages is psychologically connected just in case (a) they are psychologically similar in terms of such states as beliefs, desires, and intentions and (b) the
psychological features of the later stage are causally dependent upon the earlier stage.\textsuperscript{105}

There are two central notions in this characterization: similarity and causation. He believes that both can in principle also hold between significant others.

His reasons for choosing similarity as one of the relevant psychological features are neo-Aristotelian in nature. As he points out, on Aristotle’s view friends have similar psychological aims and live with each other while sharing thought and discussion. The shared aims that Aristotle finds important could presumably correspond with Brink’s requirement of similar intentions. The sharing of thought and discussion probably correspond with the causation requirement.

Brink backs up his similarity requirement with the following thought: He believes that we have good reason to actively seek and foster extended selves, because they can increase the complexity and diversity of the kinds of projects we can engage in. Our extended selves will pursue some projects that we also believe in, but do not have the time and capacity to execute.\textsuperscript{106} By supporting someone with similar beliefs and desires, but more time and better capacities, we can bring about what we desire without having to do everything ourselves. The idea is that anyone has a self-interested reason to seek extended selves because of this.

In a second step, once the extended selves have been found and some similarities and causal connections have been established, these relationships as such give us an additional non-instrumental reason to care about our extended selves. The reason is that the similarity we have come to share with them is just like the similarity we share with other temporal parts

\textsuperscript{105} David Brink, “Rational Egoism, Self, and Others,” p. 351.
\textsuperscript{106} Brink, “Rational Egoism, Self, and Others,” p. 355-57; Aristotle, \textit{Nicomachean Ethics}, 1169b5-6, 1170a5-7.
of our own self. We subsequently have a non-instrumental self-interested reason to be concerned about their good, Brink argues.

There is also a further important feature of the similarity requirement: It is not similarity simpliciter that matters to Brink, but continuous similarity, or in other words similarity as a transitive notion. This means: If A is similar to B and B is similar to C, then A and C will also stand in the relevant kind of relationship with each other. Even if the degree of similarity between A and C is rather low, Brink is of the opinion that such a transitive chain of similarity would be a proper basis for special concern.\(^{107}\)

The second central element in Brink’s characterization is the notion of causation. Condition (b) of Brink’s account of the person says that the psychological features of the later stage have to be causally dependent upon the earlier stage. This causation restriction implies a restriction on the “extent” of persons and their extended selves. It has the function of reining in the otherwise far-reaching implications of the account. If all person stages who are “similar in terms of such states as beliefs, desires, and intentions” are pretty much just like one person, then the account could theoretically come close to stipulating that there is one big mass self. Introducing the restriction that those beliefs, desires and intentions have to have been caused by earlier stages of the same self introduces an important qualification. The causation restriction limits the relations that matter to cases with immanent causal chains.\(^{108}\) Of course it also limits the extent to which the relations that are important can occur among separate persons or extended selves. The causation requirement seems to apply primarily in the intra-personal case or among family members and close friends.

\(^{107}\) Brink, “Rational Egoism, Self, and Others,” p. 362.

\(^{108}\) The causation relation does not have to be mutual. It is asymmetrical and may or may not occur in both directions. Compare for example Brink, “Rational Egoism and the Separateness of Persons,” pp. 126-27.
However, Brink’s causation qualification nonetheless allows the inclusion of more distant persons as extended selves—at least to a lower degree. The reason: He argues that the relevant cause may be indirect. Even complete strangers often participate in a shared social system in which they influence each other’s beliefs, desires and intentions indirectly. In fact, the view is compatible with the claim that there is a very low and very indirect degree of causal relatedness between all members of a nation or even between people on opposite sides of the planet.\(^{109}\)

In other words, just like the similarity requirement, he imagines the causation requirement to be transitive. There need not be exact similarity or direct causation between two selves in order for those to be tied together in the relevant way. If X has direct causal contact with Y and Y has direct causal contact with Z, then X and Z stand in the relevant kind of causal relationship.\(^{110}\)

### 4.3.1 Consequence: A Bigger Scope for Self-Bias and Overlap with Morality

Brink believes that the extended scope for self-interest that his account supports is able to provide a fundamental justification of morality against amoralists.\(^{111}\) Amoralists believe that they have no non-instrumental reason to be moral and should benefit themselves instead. Brink believes that the sameness of the psychological relations within an individual self and with extended selves gives everyone at least some non-instrumental reason to benefit those others.


\(^{110}\) Brink, “Rational Egoism, Self, and Others,” p. 366.

\(^{111}\) For a discussion of how this argument matters to his metaethical motivational externalism, see Brink, Moral Realism and the Foundations of Ethics, section 3.
He contrasts this extended-selves justification of morality with what he calls the “strategic defence”\textsuperscript{112} of morality. The strategic defence claims that an agent has reason to act morally because doing so is the best way of maintaining the appearance of cooperation with others that is necessary in order to receive the benefits of a social system of cooperation. The strategic defence is counter-factually instable. Whenever it is possible to maintain the appearance of cooperation without actually cooperating, it seems as if the rational agent has a supreme reason to do this rather than to actually cooperate. This way, he can secure himself the benefits of social cooperation without having to contribute.

Brink believes that a defence of morality by appeal to an account of extended selves fares much better than the strategic defence, because it is counter-factually more stable. Imagine a group of extended selves who are engaged in a shared system of cooperation. One day, one of them discovers that she could maintain the appearance of cooperation and continue sharing the benefits without actually contributing. On the strategic defence she would have supreme reason to do this. However, when considering the nature of the group members as her extended selves, she presumably does not. Since the relationship in which she stands with her extended selves is very similar to the relationship she has toward her own individual self, betraying the others would be like betraying herself.

It is possible to employ the notion of compensation to support the point. Initially, egoists think they will experience compensation if they deny themselves a present benefit in order to achieve another benefit in the future. However, they do not initially believe that a benefit to someone else, including close ones, can ever compensate them. Their beliefs about compensation are time-neutral, but agent-relative. If the egoist can get convinced, though, that he can experience compensation if a benefit goes to an extended self, this would have to

\textsuperscript{112} Brink, “Rational Egoism, Self, and Others,” p. 344-49.
change his attitude toward his close ones, even though he will not have to give up his agent-relative convictions about the possibility of compensation.

4.3.2  Brink’s Omission of First-Person Memories

The first aspect of Brink’s account that deserves a closer look is the fact that his requirements exclude something that has traditionally played an important role in personal identity theories: the continuity and connectedness of memories. Instead, Brink concerns himself with beliefs, desires and intentions. Not only the earliest explicit author about the subject, John Locke, but also contemporary writers such as Parfit and Sydney Shoemaker explicitly include memories in their accounts of personal identity. This makes the exclusion of memories rather unusual and worthy of an explanation.

The omission gains further significance in case one believes that memories possess a distinct phenomenological character that is usually only accessible “from within” a particular chain of experiences. It is often thought that qualia, the content of our sense impressions, can only be accessed by individuals themselves, not by their friends. The experience of what the difference between red and green is like for Joe is in this sense an entirely subjective fact. Maybe experience memory is just like that. Epistemic access to what exactly it is like to remember a particular event “from the inside” might be restricted to individuals. If this is the case, then individual persons will naturally be rather separate from each other.

In contrast, the psychological relations specified by Brink will sometimes apply across what we ordinarily take to be separate persons. It is perfectly plausible to say that separate persons are sometimes “psychologically similar in terms of such states as beliefs, desires and intentions” and that their mindset is causally influenced by other people. It is the fact that these relations can apply in cross-personal cases as well as in intra-personal cases that does the work in Brink’s account.
While Parfit’s reductionism seeks to shrink the gap between rational egoism and morality by *shrinking* the scope for justifiable self-bias, Brink tries to accomplish the same by *increasing* the scope for self-bias. In other words, Brink believes that the desired effect can be achieved without arguing that the egoist’s rationale is mistaken, while Parfit is trying to show that the egoist’s rationale is mistaken. If close others are part of one’s extended self, then even the self-biased have a non-instrumental reason to be interested in their good. Benefits to their extended selves would be like advantages for themselves.

Parfit takes great care to rebut accusations according to which utilitarians treat humanity as one big “mass person.” He points out that utilitarians might believe in additional separateness even within single lives rather than conflation into a group self. He thinks there is not just separateness between separate persons, but also a significant degree of separateness between the temporal stages or even moments of an individual person’s life. At first glance, Brink’s view seems to leave a lot less room for separateness. But as we shall see later, some separateness between individuals and their extended selves enters his picture through the backdoor.

When choosing to exclude memories as relevant features of what a person is, Brink seems to want to align himself with the tradition of giving supreme importance to the distinctly rational and deliberative faculties of humans. He supports this emphasis by appeal to Aristotle’s views on the rational and deliberative faculties. He also aligns himself with John Locke’s understanding of personal identity as a forensic term that sees persons as agents who
act willfully and can be held responsible for those actions. This focus on agency and responsibility likewise emphasizes deliberative and rational capacities.\(^{113}\)

However, neither Aristotle nor Locke ignores memories. For example, Locke emphasizes the importance of memory for responsibility. In his notorious passage about memories he even claims that we are never responsible for things that we have irrevocably forgotten:

> For whatsoever any substance has thought or done, which I cannot recollect, and by my consciousness make my own thought an action, it will no more belong to me, whether a part of me thought or did it, than if it had been thought or done by any other immaterial being anywhere existing.\(^{114}\)

This passage suggests that the rational and deliberative faculties might not be the only ones that matter. The phenomenological access that persons have to their own history might also be an integral part to our concept of personal identity. In fact, it might be the very thing that the concept of a person adds to the concept of an agent—the special phenomenological access to a psychological history, including events that do not turn out to influence our rational deliberations or actions.

Clearly, phenomenological experiences and recollections are important for being able to have rational and deliberative faculties. It is impossible to deliberate if we forget in one second what we thought in the previous one. But one can choose to give experiences and


recollections importance in and of themselves or only insofar as they matter for the rational and deliberative faculties. Persons have access to millions of recollections that never actually end up contributing to any rational deliberations or actions. If phenomenological experiences and recollections matter in and of themselves, then they will have to be seen as contributing to what a person is irrespective of this.

Brink can value the phenomenological only insofar as it contributes to or is inherent in the rational and deliberative faculties. He could defend his choice by arguing that the phenomenological is implicitly taken care of in the intentional states he stipulates as important, in particular in the notion of a belief. Memories are simply beliefs about the past, he could argue, and since beliefs are included in his account, so are memories. A memory of a long gone beautiful flower might be described in terms of beliefs about what exactly this flower looked and smelled like. Moreover, he could say that actively remembering the flower is a form of deliberating about it.

The importance of memory might also be included in the importance he assigns to intentions. Intentions are generally worthless if we cannot remember them at all. Most actions are not completed instantaneously as soon as the intention is formed. Some intentions will not turn into action until a specified later point in time. They also often take years to complete. This means that intentions usually have to be recalled in order to do their work. Similar things seem to be true of desires. If we forget them right away, they will hardly lead to deliberations or actions. Moreover, many of them obviously have a phenomenological component. A desire to drink, for example, will usually come with a dry feeling in the throat. In fact, the desire might not even arise without the dry feeling in the throat.

However, both the claim that memories and experiences alone have special first-person features, as well as the claim that beliefs, desires and intentions have these features
built into them seem to introduce some room for fundamental separateness between persons. If either claim is true, then extended selves will never acquire the same status as parts of individual persons. There would always remain some fundamental separateness along the lines of persons as ordinarily construed, because the phenomenological character of experiences and memories can usually not be shared across such boundaries.

It is not possible to do the concerns about memories full justice here. But it is possible to come to some conditional conclusions: If one believes that memories (and maybe other mental states as well) have a phenomenological character unlike beliefs, desires and intentions and/or cannot be shared among separate persons, and those features matter, then Brink’s account of extended selves is incomplete. Moreover, it might be incomplete in a way that exaggerates the possibilities of forming extended selves. If the phenomenological character of experiences and memories is important to what a person is, then this might well speak in favour of some fundamental separateness between individuals that keeps people from fusing entirely into larger extended selves.

Likewise, if one is convinced that beliefs, desires and intentions have recollective aspects built into them and might even have phenomenological aspects of their own, then Brink’s extended selves are to be taken with a grain of salt. In this case, there will always remain elements within Brink’s account that will divide people along traditional lines. Beliefs, desires and intentions will end up having an aspect that cannot be shared by separate persons.

There is one more consideration to address before moving on to the further features of Brink’s proposal. Traditionally one of the concerns with memory as a criterion for personal identity was that it could be circular. This is best illustrated by showing that there are essentially two possible ways to express the content of my memory of bottle breaking, which we might normally express as “I remember breaking the bottle.” If the content of this memory
contains the identification between the me now and the person who broke the bottle, then explaining identity by appeal to such memories might be circular. Consider these two ways of rephrasing the recollection:

A) “I remember I broke the bottle.”

B) “I remember a bottle was broken.”

If B) is the proper content, we still need to establish the identity between the bottle breaker and my present self. If A) is the proper content and we base our notion of personal identity on memory, then the account is circular. For if I remember that it was me who broke the bottle, then part of my memory is a belief in the identity statement in question. It contains not only the claim that there was bottle-breaking, but also the claim that my present self is identical to the bottle breaker.

However, there are a couple of ways of dealing with the circularity problem that Brink could avail himself of. One standard way of dealing with the circularity problem is to introduce the notion of quasi-memory or q-memory. Q-memories are memories without the identity requirement, for instance a memory of having broken the bottle without the jump to the conclusion that the person who did it is identical to the one who remembers it now. This would avoid the circularity of presupposing the identity statement in the content of the memory. Another possibility would be to insist that psychological persistence involves more than merely continuity of memory. It involves character traits and tacit skills, which cannot fall prey to the circularity problem of the memory criterion or to the problem of misremembering or being deceived by an evil demon.

Brink could have easily added q-memories to the psychological relations that need to be present in order for us to be able to speak of the same person. By calling them q-memories rather than memories he could have escaped the circularity charge. It can therefore not have been the reason for his omission of memories. He could have also adopted the second solution. In fact, since he gives particular importance to beliefs, desires and intentions as well as character traits, his account might have been particularly suitable for avoiding the circularity problem about memories by appealing to additional kinds of mental characteristics.

4.3.3 Brink’s Asymmetry about Discount Rates

There is also a feature of Brink’s account that seems to be inconsistent. Brink explicitly argues that both similarity and causation are to be seen as transitive notions in both intra-personal and cross-personal cases. If Tom at t₁ and Tom at t₂ stand in a relationship of similarity and causation and the same is true of Tom at t₂ and Tom at t₃, then Tom at t₁ has a non-instrumental self-interested reason to be concerned about Tom at t₃. Likewise, if A stands in a relationship of similarity and causation with B, and B stands in a relationship of similarity and causation with C, then A and C are to be seen as linked in the relevant way. A should have some non-instrumental self-interested concern for C. However, Brink also argues that there should be a discount rate for cross-personal concern, while there should not be a discount rate for intra-personal cases. This position is potentially problematic.

With regard to cross-personal cases, or cases involving extended selves, he argues that both the similarity of beliefs, desires and intentions as well as the directness of causation among extended selves are a matter of degree. The non-instrumental concern for extended selves that the egoist can justify is proportional to the degree of psychological similarity and the directness of causation. There will subsequently be a discount rate for concern for one’s

extended selves. For example, the degree of similarity and the directness of causation will be high between Sam and his close marriage partner. But they will be fairly low between selves who are merely compatriots who have never met.

With regard to intra-personal scenarios his position is much trickier. He seems to think that the degree of similarity and the directness of causation are always much higher in the intra-personal case. While this could in fact turn out to be an unproblematic empirical consequence of his view, he also pushes the point further by arguing that there is a principled reason for this: namely that there is no discount rate in the intra-personal case and that it is plausible for the egoist to be equally concerned about all parts of his own existence. Unlike concern for continuous extended selves, concern for continuous parts of one’s individual self is therefore not a matter of degree.

This asymmetrical way of dealing with intra-personal and cross-personal cases is rather surprising. For example, it seems to have a detrimental impact on the dialectic goal of his metaphysical egoism. The goal of his metaphysical egoism is to show that even egoists will usually have reason to do what morality demands because their relationship toward others is like the relationship they have toward themselves. His endorsement of a self-other asymmetry about discount rates seems to endanger this central feature of his account to a significant extent. Egoists could now generally have a stronger reason to benefit their own future selves rather than their extended selves. Similarity and causation with extended selves might only ever reach partial degrees, while it would remain continuously high within the egoist’s own life. Some tension between egoism and morality would therefore always remain.

\[\text{117 Brink, “Rational Egoism and the Separateness of Persons,” p. 128.}\]
\[\text{118 Brink, “Rational Egoism and the Separateness of Persons,” section 8.}\]
Brink offers two reasons for his asymmetry about discount rates. As he correctly admits in the Postscript to “Rational Egoism and the Separateness of Persons,”\textsuperscript{119} his first argument can be seen as inconsistent and cannot stand on its own. This argument rejected the discount rate in the intra-personal case on the ground that even the temporally most distant parts of individual persons are still equally continuous with each other. This means that he treated continuity as an all-or-nothing relation. At the same time, he endorsed the discount rate in the cross personal case on the ground that the degree of continuity can decrease. It is inconsistent to treat continuity as an all-or-nothing-relation in one case, but as a matter of degree in the other.

The Postscript subsequently offers a second argument for the self-other asymmetry about discount rates. This argument depends on what he takes to be the proper understanding of the relationship between parts and wholes of persons. The case that is supposed to speak in favour of an intra-personal discount rate is the one in which the distant temporal parts of a person are only loosely connected with each other. But this case, he argues, “demonstrates only a fact about the relation among the parts of my life, not a fact about the relation between the parts and me.”\textsuperscript{120} In the intra-personal case, all temporal parts are equally parts of a whole, he believes.

But why should networks of extended selves not possess some wholeness status of their own? Why should it not be the case that there are individual wholes and extended wholes? In fact, how does he even want to distinguish between intra-personal and cross-personal relations on the basis of an account in which the relevant relations can occur in both cases? The asymmetrical position only makes sense if one believes that individual persons

\textsuperscript{119} Brink, “Rational Egoism and the Separateness of Persons,” pp. 133-34.
\textsuperscript{120} Brink, “Rational Egoism and the Separateness of Persons,” p. 121.
possess some kind of wholeness status and networks of extended selves do not. If both or neither possessed a wholeness status, then the treatment of intra-personal and cross-personal cases would have to be symmetrical. If only networks of extended selves possessed wholeness status, then the account could not make any special claims for individual selves.

Brink’s stipulation of a wholeness status for individuals seems to simply presuppose the truth of anti-reductionism about individual persons. It would mean that there is some kind of metaphysical unity or identity relation that grants individual separate persons their privileged status. But this position is not actually open to him since he argues that his conclusions are true even if reductionism about personal identity is true.\(^\text{121}\)

Furthermore, this argument seems to assume that identity matters for practical purposes. He assumes that it should result in the lack of a discount rate for intra-personal concern. At the same time, though, Brink regularly makes extensive use of a more reductionist Parfitian view of what matters in order to support his account of extended selves. For example, he insists that in the hypothetical case of branching what matters remains preserved. Since identity is not preserved in branching, the case implies that identity does not matter. He believes this to be the strongest case in which someone could have a non-instrumental self-interested attitude for someone with whom he or she is not identical, just like in the cases of real life extended selves. This argument assumes that identity does not matter. It is inconsistent to appeal to a view in which identity of individual persons matters in some cases, but not in others.

Moreover, as pointed out above, if identity or wholeness of individual persons does matter and networks of extended selves do not have such a status, then Brink’s conclusions

\(^{121}\) Brink, “Rational Egoism and the Separateness of Persons,” p. 129: “Though the egoist rationale does appeal to the separateness of persons, it is not undermined by reductionist accounts of personal identity.”
about the egoist’s reason to benefit extended selves are endangered. The reason is that the egoist can always maintain that of all the things that do matter, identity applies only among the parts of his own life. There will therefore always be a principled reason to be more concerned about the various parts of one’s own life. Even if beliefs, desires and intentions that are shared with others would also give the egoist some reason for concern, this might not turn out to be an *overriding* reason, because the wholeness relation is lacking for networks of extended selves.

For example, imagine Brink’s egoist had to decide between benefiting his own self thirty years from now or the present self of a friend with whom he shares just as many beliefs, desires and intentions. Brink’s egoist would actually have *some* reason to benefit his friend, because of the shared beliefs, desires and intentions. But he would have *overriding* reason to benefit himself. The reason is that the wholeness relation and the lack of a discount rate for the parts of his own existence always give him additional reason to benefit himself instead of his friend.

Nonetheless, there are some cases in which Brink’s egoist might have overriding reason to benefit his friend. Imagine the size of the benefit that the friend could get would be significantly larger than the size of the benefit that would go to his own future self. In this case, the extra size might in fact outweigh the additional benefit that comes from giving the benefit to a self that one stands in a wholeness relation with. This means the egoist’s calculation could sometimes tip in favour of extended selves. But in any case the extended selves will always have to compete with the additional benefit that the egoist believes to come from benefiting his own self. This means that Brink’s account will sometimes—but not always—give an egoist overriding reason to benefit his friends. This is what Brink’s account accomplishes.
Nonetheless, this conditional success does not solve the problem that it is inconsistent to argue in favour of asymmetrical discount rates in the way Brink does, while at the same time maintaining that his account is true irrespective of whether reductionism about identity is true. Of course the claim that the relevant psychological relations hold to a stronger degree within persons in the ordinary sense is a matter of common sense, as Sidgwick argued. Probably most people’s intuitions go in this direction. But the way in which Brink tries to explain this seems inconsistent.

Two possible ways of avoiding the inconsistency suggest themselves. First, he could drop the asymmetry and endorse discount rates for both intra-personal and cross-personal cases. He could then add the empirical claim that continuity of beliefs, desires and intentions will most often turn out to be stronger in intra-personal cases. He would subsequently have to hope that empirical investigations would confirm his claim. The other possibility for achieving an asymmetry without unwittingly endorsing the truth of anti-reductionism would be to include the phenomenological character of memories and experiences among the relevant psychological relations. They will never hold between extended selves and would therefore imply a privileged status for individuals. When formulated in terms of q-memories this solution would not be circular and would not assume the truth of anti-reductionism.

Moreover, he could argue that it is at least in principle possible for there to be a continuous chain of phenomenological recollections throughout the entire existence of a person—a possibility that cannot occur in the cross-personal case. This possibility could be reason to refrain from a discount rate for phenomenological self-relatedness, while continuity of intra-personal beliefs, desires and intentions could still come with a discount rate. Parfit believes that it is the actual degree to which people are able to recall that matters. He also believes this degree can be low or even zero. But Brink could insist on q-phenomenology as
an all-or-nothing relation, together with continuity of beliefs, desires and intentions as a matter of degree. This would achieve the exact same as his denial of a discount rate for intra-personal relations—without running into an inconsistent asymmetry.

4.3.4 Egoism, Self-referential Altruism, and Morality

Brink’s account, if successful, leads to significant overlap between egoism and morality. We may still ask whether his self-extended egoism would always yield the exact same results as morality though. Brink’s account relies on the idea that actions are over-determined when it comes to the reasons one could have for committing them. In addition to other-regarding reasons for benefiting others, there are generally also self-referential reasons in favour of the very same action. This makes it the case that complying with the other-regarding demands of morality cannot be seen as irrational from a prudential perspective.

For example, the other-regarding reasons of morality demand that we support our friends in need. But a self-extended egoist who is never moved by altruistic considerations would also have reason to support a friend in need. After all, the friend is the egoist’s extended self. This means that the amoral egoist cannot claim that this demand of morality would be fundamentally contrary to the commands of rational self-interest or prudence, even if non-compliance with the demand would never get detected. In fact, self-interest and morality will both require helping a friend in need.

Depending on the moral theory one endorses, morality and self-extended egoism might even recommend the exact same actions. This would be the case if the moral theory one endorses acknowledges the exact same inter-personal discount rate for other-regarding moral obligations as the one that self-extended egoism acknowledges for the self-interested reasons. For example, morality might demand that we care x times more about persons who are close to us than about others. If the degree of psychological relatedness that we share with our close
ones is also x times higher than our relatedness with others, then morality and prudence would have to demand the very same actions.

Nonetheless, there might still be cases in which morality and self-interest do not demand the exact same degree of concern for someone. For example, morality might demand other-regarding actions that give weight to others over-proportionally to the degree of our psychological relatedness to them. In such a case, the amoralist would not necessarily be willing to act according to the morally required degree of concern. This would in particular be the case if the action that morality demands would be in conflict with another action that would bring more benefit to the amoralist or his closer extended selves.

Overall, this means that self-interested reasons and moral reasons might not always correlate in strength, but an amoral egoist will always have at least some non-instrumental reason to do what is morally required. Moreover, irrespective of the kind of moral theory endorsed, the view would be counterfactually more stable than the strategic defence, because betraying a friend amounts to betraying oneself. Any potential dualism of practical reason seems subsequently less fundamental than Sidgwick thought.

However, before coming to terms with this underlying motivation for Brink’s self-extended egoism, it is important to consider some potential objections. For example, take a case in which we have some reason to benefit an extended self, but also a stronger, overriding reason to benefit ourselves. One objection would insist that acting on the weaker reason is irrational. If this is correct, then Brink’s rationale does not really work out. He wants to be able to say that acting on behalf of significant others is at least never irrational, even if there is an overriding reason to benefit oneself. But maybe following weak and overridden reasons is irrational. Brink would simply have to deny this and insist that following an overridden reason is rational.
Another objection someone could raise is that self-extended egoism does not serve its purpose in the right kind of way. The reason the self-interested egoist would have to engage in the moral action would not be the kind of reason that morality would require him to have. Morality requires that we care about others for the sake of them as others. Extended egoists care for their extended selves for the sake of themselves. They do not grant those others the status of entirely separate beings. Morality might require, however, that we care about the recipient of a benefit in his or her capacity as another, separate entity.

One response that Brink could give would be that he treats extended selves not just as selves, but also as others. This seems certainly be true of passages in which he explains why persons have a reason to actively seek and cultivate other selves in the first place. In these passages he argues that having other selves “allows me to engage in more complex and more diverse activities than I could on my own.”\textsuperscript{122} The idea is that people can usually exercise their interests in various ways and have to choose and refrain from pursuing some of the options due to restrictions of time, resources and capacities. When cultivating other selves, though, they will be able to have more such projects in their lives by watching and discussing them. In this context Brink also puts forth some arguments that emphasize the importance of the dissimilar and causally independent parts of one’s extended selves. He writes,

\begin{quote}
my other-self is not a mere clone of me. Clearly, I will diversify my experience more by interacting with someone who has somewhat different interests and experiences.
\end{quote}

\begin{flushright}
\textsuperscript{122} Brink, “Rational Egoism, Self, and Others,” p. 355-57; Aristotle, \textit{NE} 1169b5-6, 1170a5-7.
\end{flushright}
And cooperation in complex projects will often be enhanced when participants have different strengths and talents.\textsuperscript{123}

These passages suggest that the rationality of caring about others does not just extend to those aspects of other persons that are psychologically similar to oneself and stand in a causal relationship to oneself. In fact, we should also care about these dissimilar parts because they are the ones that provide even more chances for the extended exercise of our own interests.

Admittedly, the rationality of interest in the dissimilar parts of others has an instrumental ring, but it also suggests that the interest in extended selves does not just happen insofar as those others are \textit{me or like me}. The regard also seems to be other-regarding in the sense of caring about dissimilar aspects. In fact, the expansion of our opportunities in life through fostering extended selves gives us reason to care about others irrespective of whether they are already part of our extended self. Our self-interested care about them does therefore not imply a complete denial of their status as others.

The observation that there is some genuine other-regarding concern in Brink’s picture of egoism raises the question why Brink insists on the label egoism. In fact, C. D. Broad’s “self-referential altruism” seems to be extensionally equivalent to Brink’s self-extended egoism—even though it is called \textit{altruism}. Broad distinguishes between various ways in which an action can be self-related. By a self-referential action he means an action in which “the relationship of the other person or thing to oneself acts as a strong egoistic motive-stimulant, but one’s primary desire is that the other person or thing shall be in a certain

\textsuperscript{123}Brink, “Rational Egoism, Self, and Others,” p. 356.
Such actions require both a non-instrumental interest in the other for his or her own sake, but the relationship within which one stands with the recipient provides a self-referential element.

Brink’s view presents the same picture: The interest that the amoralist takes in his extended selves is non-instrumental, because he cares for them for their own sake. The reason he cares for them for their own sake consists initially in their capacities that allow the expansion of his opportunities and later also in the psychological relationship of similarity and causation that makes them his extended selves. His view even grants the extended selves a limited status as others in respect to those parts of their psychologies that do not stand in the relevant relationship of similarity and causation to him.

Broad was searching for the kind of view that common sense judgments implicitly endorse when developing his self-referential altruism. When discussing his notion of the self-referential, Broad is investigating whether all motives for action are egoistic in the sense of having a self-referential element or whether there might not also be motives that are genuinely other-regarding without being also self-referential. In this sense, Broad’s investigation has the opposite upshot of Brink’s. Broad wants to show that there are actions of which we can say that they have a genuinely other-regarding motive that is not also self-referential. Brink, on the other hand, wants to show that all other-regarding actions also have a self-referential aspect.  

Despite this difference, though, their conclusions seem to be compatible. Broad expects that actions can sometimes be motivated by entirely other-directed motives, for

---

125 For a related discussion about the justificatory order between self-concern and other-regarding concern see Whiting, “Impersonal Friends,” pp. 4-6.
example if a young priest were to decide to help a number of unknown lepers in Asia. However, he also admits that we will never know whether the other-regarding motive of this action was the one and only one at work. We can never know whether desires for heroism, glory and a saintly reputation were not in and of themselves a sufficient motivation for the priest. We cannot prove that there was ever a person who committed such a saintly act and who was absolutely certainly not moved by a self-directed desire for saintliness.

Brink would have to be happy to subscribe to the same conclusion. If we cannot prove that there is ever an act in which self-directed motives play no role at all, then his attempt at convincing the amoralist that he has always at least some reason to do what morality commands cannot be proven to fail. Even a saintly endeavour like living among lepers may have some self-directed motivation. This means that it is at least not entirely and fundamentally contrary to reason to do such a thing.

4.4 A Hybrid Account about Overlap between Self-Bias and Other-Regarding Concern

Despite their differences in focus, the outcomes of both Brink’s and Parfit’s accounts depend on the relative importance of intra-personal and cross-personal relations. When deciding about the relative importance of intra-personal and cross-personal relations we could consider a large number of different possibilities. I will focus on three, two of which have been anticipated either by Parfit or both Parfit and Brink. The third one will be a further proposal that has not been considered by them. According to the first option, one could argue that as the intra-personal relations fade, both intra-personal and cross-personal relations can become equally strong and important. Second, one could argue that intra-personal relations are always

126 Various authors have raised this point, for example Broad, “Egoism as a Theory of Human Motives,” pp. 255-58.
somewhat stronger and more important. Third, I will discuss the possibility that as the intra-
personal relations fade, cross-personal relations might actually become stronger and more
important.

1.
Let’s look at what the first option would yield, the option that cross-personal psychological
relations and intra-personal psychological relations can in some cases become equally strong.
There is some evidence that Parfit holds this view. For example, in section 114 of Reasons
and Persons he argues, “[a] reductionist is more likely to regard this child’s relation to his
adult self as being like a relation to a different person” (emphasis added). This passage
seems to imply that the psychological relations between remote parts of an individual self and
two separate persons can be of equal strength and importance.

This position depends on the idea that the discount rate for intra-personal
connectedness can bring the ground for special care about one’s own future selves either all
the way to zero or at least down to the very same level to which we have reason to care about
other people’s selves. The idea is that intra-personal relatedness—including
phenomenological experience memory—tends to fade over time. The relatedness between
myself now and myself in the distant future might in fact be so faint that the relatedness
between myself now and the self of a separate person can become equally strong.

This implies the following. On a short-term basis it might still be possible for self-bias
to arise and be justifiable. During periods of strong intra-personal relatedness, the relatedness
with other people will be less strong. This ground for self-bias is not of the same nature as
rational self-interest as it is ordinarily construed. The reason is that it would not hold equally

127 Parfit, Reasons and Persons, p. 335.
between all temporal parts of one’s existence, as stipulated by Sidgwick’s requirement. Such self-bias would generally only be plausible within a limited period. For example, at the age of 44 this time period might extend five years into the past and five years into the future. At the age of 48, the time period would then likewise extend five years into the past and five years into the future. Within this forward moving time frame self-bias might be justifiable due to the particularly strong intra-personal connections that hold within it.

However, the relationship between one’s present self and a future self outside this time frame might be weak enough so that the relationship between one’s present self and the self of a separate person is just as strong. This would have the following further implications. When deciding whether a particular benefit should go to oneself one year from now or to the present self of another person, there would be some room for rational self-bias and the choice that the benefit should go to oneself. On the other hand, when deciding whether a particular benefit should go to oneself twenty years from now or to the present self of another person with whom we are or will be just as strongly connected as with ourselves twenty years from now, there would be less no room for self-bias left.

For example, imagine Mr. X had to decide whether to benefit his own self twenty years from now or the current self of his dear wife. The relationship between X’s current self and his own self twenty years from now might be so weak that the relationship between his current self and the current self of his wife is just as strong. This means that X has just as much reason to benefit the current self of his wife as his own self twenty years from now. There is no rational room for self-bias anymore in the decision X is facing.

Of course the size of the expected benefits would be relevant in addition to psychological relatedness. Imagine X had to decide between benefiting his own self six years from now or the present self of his wife. Presumably within the time frame of six years there
would still be some room for self-preference left, albeit to a reduced degree. But if the benefit that could go to the wife would be a lot larger than the benefit that his own self in six years would get, then balance might actually tip in favour of the wife. The small amount of self-bias that would still be plausible in this case can arguably get outweighed by the large size of the benefit that could go to his wife with whom he is only slightly less connected than with his own self six years from now.

It is important to note that the view under consideration requires us to make a further choice. We will have to decide whether to treat our relatedness to all others as always equal or whether our relatedness to others comes with its own discount rate based on similarity and causation as Brink suggests and Parfit very briefly hints at. The former view would give us reason for equal consideration of all other persons. If we had to decide whether a certain benefit should go to one other person rather than another, we would have to base the decision on something like the size of the benefit. In any case, the relation within which that person stands to us could play no role. The latter view, in contrast, would allow us to give some preference to those others we are close with. In addition to just paying attention to the size of the expected benefit, we would be able to give preference to persons we are close with.

There are further questions to be asked. For example, we may ask whether there should be a further discount rate for the relatedness between X’s present self and the increasingly remote future selves of his wife. The question might not be whether X should give the benefit to his wife’s present self, but rather to her future self. If the choice is between X’s own self twenty years from now and his wife’s self twenty years from now, it seems unlikely that the basis for self-bias would disappear, given the stipulation that self-regarding and other-regarding concern can at the most become equal. If the choice is between X’s own self twenty years from now and his wife’s self three years from now, then there might actually
be no room for self-bias in favour of X’s own self. The reason to benefit the remote part of himself might be so low that the reason to benefit his wife three years from now is just as high.

I will not go into further detail about these matters, but it is worth spelling out once more what a view according to which intra-personal relatedness and cross-personal relatedness can sometimes become equally strong would imply for the tension between egoism and morality. It would mean that courses of action that benefit one’s own future self would remain prudentially plausible within the time frame within we are strongly psychologically related to ourselves. If the benefit would only arrive after the end of this time frame, a course of action that would benefit someone else could be equally prudential.

2.

According to the second alternative, intra-personal relations are always at least somewhat stronger and more important than cross-personal ones. Both Parfit and Brink have anticipated this alternative. In section 115 of Reasons and Persons Parfit writes: “Those claims [about compensation] treat weakly connected parts of one life as, in some respects, or to some degree, like different lives” (addition and emphasis added).\(^\text{128}\) This formulation does not say that the weak connections between remote parts of one’s own life are as weak as the connections between one’s own life and other lives, although other passages of the book leave this possibility open. Instead, it says that weakly connected parts of one’s own life are only to some degree like connections to other lives. The idea seems to be that the similarity between intra-personal and cross-personal relations can become close, but intra-personal relations will always be somewhat stronger and more important. Brink seems to have had a similar option

\(^\text{128}\) Parfit, Reasons and Persons, p. 337.
in mind. There are passages in which he argues that there is a discount rate for our interest in extended selves, but never for the various temporal parts of our own self.¹²⁹

Despite the fact that Parfit believes in an intra-personal discount rate and Brink does not, these views would have the same implication for the relation between intra-personal and cross-personal concern. Both of these views hold that the ground for intra-personal concern is generally stronger than for cross-personal concern. This conviction will make this option more intuitive than option one for people who believe that one has always at least somewhat more reason to be concerned about oneself than about others.

In terms of easing the tension between egoism and morality, this option does somewhat less work than option one, though. If intra-personal connectedness is always stronger than any connections across people, then the egoist’s rationale and morality remain to a larger extent at odds. The idea that there is always somewhat more intra-personal psychological relatedness will presumably be a sufficient reason for egoists to have a general attitude of self-bias in all cases.

Nonetheless, the strength of this egoist rationale might often be weak. One’s intra-personal relationship to one’s own distant self can be so weak that it is almost as low as our relationship toward other people. This means: When deciding whether a particular benefit should go to one’s own self or the self of another person, the calculation may tip in favour of the other person. This could be the case whenever the benefit that that person could receive is significantly larger than the competing benefit for oneself. However, if the benefits were equal, egoists acting according to this theory would see overriding reason to benefit themselves.

¹²⁹ Brink, “Rational Egoism and the Separateness of Persons,” Postscript; section 8; pp. 121, 128, 133-34.
3.

Finally, the third alternative would be that the cross-personal relations could in some cases actually become stronger and prudentially more important than the intra-personal relations, even though the intra-personal relations will usually still prevail within a time frame of, say, five years into the past and five years into the future. Neither Brink nor Parfit have anticipated this alternative in the form in which I will suggest it here. Brink dismisses relations that can in principle only hold in the intra-personal case altogether and he rejects a discount rate for intra-personal relations. Parfit raises the possibility that cross-personal relations might also matter, but does not consider the possibility that they can become stronger and more important than intra-personal relations in ordinary cases.130

This view would claim that there is room for self-bias within the time period within which we are strongly psychologically connected to ourselves, for example within ten-year units that reach five years into the past and five years into the future. With time, however, the vividness of these intra-personal relations fades. Outside the ten-year units the psychological relations that we share with others can therefore become more important, not just equally important.

For example, compare the psychological relatedness between one’s own current ten-year unit and oneself in the distant future with the psychological relatedness between one’s current unit and the current unit of a close friend. The distinctly intra-personal phenomenological character of the relations between one’s current unit and one’s unit in the distant future will have faded considerably. Most people hardly remember the exact details of their daily affairs twenty years after they took place. Especially the phenomenological character of such recollections seems to fade to a considerable extent. There will also be a lot

of external causal influences that will have shaped one’s self in the distant future. In the time until then, we will have had many encounters with other people that will have causally contributed to our mindsets—maybe even a lot more than our by then remote current self. Our degree of similarity and the directness of causation with the current selfhood unit of one’s close friend might therefore be very high. Over the course of the last five years we might have gone through countless experiences together during which we mutually shaped our belief sets as well as our desires and plans for the future. Our psychological similarity as well as the causal influence we have had on each other during our current units might be extremely high.\(^{131}\)

The first option claimed that the cross-personal relations could in this case become equally strong and important as the intra-personal relations. But it seems to me as if they can sometimes be \textit{stronger} and \textit{more} important than the faint recollections of current times that one will have twenty years from now. Why could my psychological relatedness to a close friend during our current units not be much stronger and more important than my relatedness to an entity thirty years from now with which I might not even share basic convictions anymore?

Such a position seems to have the advantage of granting the egoist that there is something special about first-person relatedness, especially phenomenological self-relatedness. This is a conviction that egoists feel very strongly about and might not be willing to give up. Moreover, the view preserves the possibility of arguing that there is a non-

\(^{131}\) It seems important to include a part of the \textit{past} within the current selfhood units of the two persons. If we had to decide at this very moment between benefiting one’s own self twenty years from now and the \textit{present moment} of another person, the decision might not turn out the same way as if we include a part of the past in our consideration. The reason is that only very few causal relations will hold between one’s own \textit{present moment} and the \textit{present moment} of a significant other. The ground for self-referential concern about significant others is not only based on present similarity, but also on causal facts about the past. See Hurka, \textit{Virtue, Vice, and Value}, p. 202: “In all their forms, the virtues of loyalty are a response to facts about the past.”
instrumental self-referential reason to be interested in other people’s good: the reason is that having done so will be better for one’s own self as well as for one’s extended selves. Apart from phenomenological experiences, psychological relations to close others can become like relations to oneself. And since the intra-personal relations tend to fade, those cross-personal relations might often be more important for oneself.

It may be argued that Parfit almost anticipated this option after all. Consider his branching case. It may well be said that the pre-branching person has a stronger reason to care about the non-identical post-branching selves than about his own self thirty years from now had he not branched. This means Parfit has anticipated that one might have a stronger reason to benefit the self of another person rather than one’s own self in the distant future.

There are two differences between the proposed hybrid approach and what Parfit’s branching case implies, though. First, comparing one’s attitude toward a potential branching off-shoot with one’s attitude toward a distant future self after an ordinary development amounts to comparing attitudes toward of two future selves, the future post-branching beings and the person’s own self thirty years from now. I am arguing that a choice between benefiting the present self of another person and one’s own self thirty year’s from now should turn out in favour of the other person. Moreover, and maybe more importantly, the branching incident involves an abnormal cause. The relevant consequence from the hybrid approach will prevail though if we add a restriction of normal causes. If we insist that both the distinctly intra-personal relations such as first-person memories as well as the shareable relations such as beliefs, desires and intentions are only prudentially relevant if they have a normal cause, the hybrid approach will remain unaffected. It poses no problem for the hybrid approach to treat cases of abnormal causes, such as branching, as cases in which a person has ended and
another one came into existence. Likewise, coming to share beliefs, desires and intentions through interaction with significant others will qualify as a normal cause.

In the quest for reducing the tension between egoism and morality the hybrid view does more work than the other two alternatives discussed. Alternative one suggested that intra-personal relations can become as weak as cross-personal relations. In those cases the egoist has equally strong reason to benefit his individual or his extended self. In all other cases, the discount rate makes sure that the plausibility of bias in favour of one’s individual self is smaller than on Sidgwick’s view, but it remains nonetheless larger than the self-referential bias in favour of others. Alternative two did worse. It suggested that intra-personal relations are always more important than cross-personal ones. According to this alternative, the plausibility of self-bias and any subsequent tension with morality would decline due to the discount rate, but generally remain stronger than the self-referential reason to benefit others. On both of these alternatives egoists will have some reason to benefit their extended selves. However, on both views egoists will only have overriding reason to benefit others if the benefits that could go to them would be larger than the ones that they could secure for themselves. The proposed hybrid account gives egoists overriding reason to benefit others even in cases in which the expected benefits are equal.

This hybrid view also seems to be in accordance with common sense. Common sense seems to hold that there can be something very tempting about securing advantages to oneself in the short term, but that it will usually not pay off in the long run. Common sense also seems to hold that the reason selfish behavior will not pay off in the long run lies in the danger of getting detected as selfish. The hybrid proposal adds a non-instrumental reason to this: The reason it does not pay off is that the first-person character of experiencing selfish advantages tends to fade over time while the good that comes to oneself and one’s extended selves from
shared and mutually caused beliefs, desires and intentions is very strong and has the potential of remaining strong for longer.

There is also a further way in which the hybrid account is in accordance with common sense. By insisting in a consistent way that both intra-personal and cross-personal relations can sometimes provide overriding self-referential reason for action, it makes no pretense that the label egoism for prudential reason is any more justified than the label altruism. If intra-personal and cross-personal relations are similar in nature, then so are egoism and self-referential altruism.

The suggested view is hybrid in the sense that it claims both intra-personal and cross-personal self-referential relations can sometimes provide overriding prudential reason for action. The hybrid account gives some limited plausibility to intra-personal self-bias according to a Parfitian discount rate. It accepts Brink’s worked out proposal and Parfit’s hint of a discount rate for self-referential interest in others. It departs from both views in holding that cross-personal relations can in some cases be more important than intra-personal ones. In this way the hybrid view is the only view that gives anyone a general overriding reason to benefit others even if the benefit that could go to one’s own self instead would be equally large.
5 Deontology and the Separateness of Persons
Both Robert Nozick and John Rawls have used the separateness of persons to argue in favour of their own view. Nozick argues that the separateness of persons supports rights as side constraints, specifically, his libertarian side constraints. Rawls argues that the separateness of persons speaks in favour of his contractually derived principles of justice. The fact that their views are fundamentally opposed in terms of the status of property rights and distributive justice makes their shared appeal to the separateness of persons rather puzzling and worthy of an investigation.

For Nozick the separateness of persons supports the idea that utilitarianism needs to be fixed with deontological side constraints. He distinguishes between an end-goal structure and a side constraint structure of individual rights. If the protection of individual rights is an end goal, then this could give reason to violate that very right if by doing so more extensive violations of the same right could be prevented. For example, if one could prevent the violation of five people’s rights by violating the same right of a single person, the end-goal structure would recommend doing so. The reason is that it would result in the most extensive rights-protection overall—at the expense of the one individual. This, Nozick argues, should be avoided. Hence, rights-protection should have the structure of side constraints, which forbid the violation of such a right as a means to any end goal, including the goal of protecting the rights of a larger number of persons.  

Nozick further insists that there can be an argument from moral form to moral content. Not only can the separateness of persons support a particular structure of rights, namely the structure of side constraints, but it can also support a particular individual right or form of liberty: his libertarian prohibition against aggression. He argues:

This root idea, namely that there are different individuals with separate lives and so no one may be sacrificed for others, underlies the existence of moral side constraints, but it also, I believe, leads to a libertarian side constraint that prohibits aggression against another.\(^{134}\)

A right against aggression is a distinctive feature of Nozick’s account. However, an even more important feature of his account consists in the claim that we ought to view property rights as libertarian constraints as well. To make a full argument for his entire view on the basis of the separateness of persons, he therefore has to establish that a property-related side constraint also follows from moral form, for example by arguing that it follows as a further consequence from the right against aggression or by arguing that it follows from Lockean (separate) self-ownership rights. If there are sufficient arguments to this effect, then the separateness of persons could support one of the most central features of a libertarian rights view.

Rawls’s appeal to the separateness of persons is also well known. In part 1 of *A Theory of Justice* he argues, “the correct regulative principle for anything depends on the nature of that thing.”\(^{135}\) Later he specifies this claim with the following point in favour of his contractually derived principles of justice:

… a love of mankind that wishes to preserve the distinction of persons, to recognize the separateness of life and experience, will use the two principles of justice to

\(^{134}\) Nozick, *Anarchy, State, and Utopia*, p. 33.

determine its aims when the many goods it cherishes are in opposition. This is simply to say that this love is guided by what individuals themselves would consent to in a fair initial situation which gives them equal representation as moral persons.\textsuperscript{136}

Rawls argues that recognizing the separateness of persons will lead to his two principles of justice. But he also informs us that it will indicate what persons will consent to in a fair initial situation. The support for the two principles of justice apparently depends on the support for contractualism. The idea seems to be that the separateness of persons speaks in favour of contractualism, and in a fair contractualist situation persons will choose the Rawlsian principles of justice.\textsuperscript{137}

This means we might be dealing with a two-step argument in Rawls’s case. First, the separateness of persons would support the contractualist procedure. This procedure could theoretically yield any principle, including the utilitarian principle. In the second step of the argument, which may or may not itself receive support from the separateness of persons, Rawls gives reasons to believe that the procedure should yield his principles of justice. Since at least the first principle, demanding basic equal liberties for all, will be viewed as a deontological principle, the argument as a whole leads to a deontological position. It is of course less clear whether the second principle is deontological. Demands such as equal opportunities and priority for the worst off seem to be consequentialist. But as long as the argument also speaks in favour of basic equal liberties, we can say that this part of it is an argument in favour of deontology.

\textsuperscript{137} This reading of the passage corresponds to McKerlie’s reading in “Egalitarianism and the Separateness of Persons,” p. 214.
What could it mean that two authors as opposed as Nozick and Rawls have appealed to the same fact to support their arguments? While both of them make use of deontological or Kantian elements in their theories, many opposed aspects of their views such as the role of property ownership, distributive justice and how extensive a state we need have given rise to endless controversies among their followers. While Nozick views property rights as deontological side constraints that we may not violate, Rawls does not extend his deontological position this far and effectively decides that we may take from the rich whenever his egalitarian distribution principle demands so. It seems rather strange that the deontological elements of both accounts are nonetheless supposed to derive equal support from the same fact about persons.

Two hypotheses seem to suggest themselves initially. First, the separateness of persons might be so indeterminate that it actually speaks in favour of both (or at least not against either) of their views. In fact, it might only speak against a common rival theory that both Rawls and Nozick have to discredit, for example utilitarianism. Second, one of them might actually respect this separateness more than the other, and can subsequently use the point more strongly in favour of his view.

These two options can also be described by using a distinction that has now been suggested several times with regard to the appeals to separateness by the two authors: the distinction between a negative and a positive argument. If the appeal to the separateness of persons is to be understood as a negative argument, then it might only speak against further rival theories, such as utilitarianism. In other words, it might speak in favour of features that

\[138\]

McKerlie seems to have first raised such a distinction in this context in “Egalitarianism and the Separateness of Persons.” It has been picked up by other authors now, for example by Brink in “The Separateness of Persons, Distributive Norms, and Moral Theory” and by Zwolinski in “The Separateness of Persons and Liberal Theory.”
both Nozick’s and Rawls’s views share, but that are not featured by utilitarianism. This would only rule out the competitor, but would not tip the scale in favour of either Nozick or Rawls.

On the other hand, if the appeal to separateness is to be understood as a positive argument, it might have to lend conclusive support to either Nozick’s rights view or Rawls’s egalitarianism. One single issue, clearly stated, should not be able to lend conclusive positive support to two fundamentally opposed theories. If the appeal to separateness is a positive argument in this sense, it is therefore unlikely that both Nozick and Rawls can derive equally conclusive support from the separateness of persons. One of them should be able to claim that he respects the separateness of persons more than the other.

For the sake of completeness, it should be pointed out that the appeal to separateness might also turn out to be both a negative and a positive argument. It might speak against shared rivals, while at the same time supporting a positive view. If this is true, then both Nozick and Rawls would have some claim to be right. The one who could not count the positive argument in his favour could at least still say that the negative argument is the one that was intended and that it works.

Of course there might be other explanations as well. For example, Rawls and Nozick might have different properties of persons in mind when referring to the separateness of persons. This would complicate matters. In this case, each of them could be right in that his favourite property leads to his respective position; we would then have to figure out whose favourite property is more important. If both properties are equally important, we are faced with a dilemma. The possibility of such a hidden equivocation between the two authors demands special care when investigating what exactly they both mean by the separateness of persons.
In the following investigation I will analyze the arguments by Nozick and Rawls and try to determine what, if anything, the appeals to separateness can really establish on their behalf. I will investigate whether their arguments can be seen as positive arguments in favour of libertarian side constraints or contractually derived principles of justice and also consider whether they can yield a negative argument against rival theories.

5.1 How Separateness might Support Deontological Side Constraints

Nozick’s argument in favour of side constraints on how we may treat people proceeds partly by attacking the competing end-goal structure of rights. The feature of the end-goal structure that Nozick seeks to discredit is the one that tells us to violate someone’s rights whenever this will prevent a larger number of violations. His side constraints, on the other hand, are supposed to represent prohibitions on how we may act rather than prescriptions of an end-goal.

At this point someone could obviously object that one might respect the separateness of persons more by protecting more persons from rights infringements. In fact, Nozick himself raises this question by asking whether the side constraint structure might not have to be seen as irrational in this respect. If respect for the separateness of persons gives persons a moral status that prohibits us from imposing rights infringements on them, then why would this respect not demand that this should happen to as few persons as possible whenever we cannot prevent it entirely?

Nozick seeks to support his position by using some familiar means-end terminology. The idea seems to be that individuals who possess a certain moral status may never be used as a mere means for the ends of other people. As he argues, aiming at never treating people as a

---

means at all might be too stringent a condition that would seriously curtail the possibility of bilateral exchange. If we could never treat persons as a means at all, there could hardly be any contracts. In a contract, we usually treat the other party at least partly as a means to getting what we want. Aligning himself with Kant, Nozick believes that it is acceptable to treat someone as a means, if we also treat them as an end at the same time.

In a next step, Nozick ties these means-end considerations to his argument in favour of side constraints. He argues: If we were to violate the rights of a particular individual in order to prevent similar rights infringements for various other persons, we would be using this individual as a mere means for our goal of benefiting the other persons. This is to be avoided and Nozick insists it can only be avoided by endorsing side constraints.

At this point, one may ask whether it is actually true that violating the rights of a particular individual to prevent similar rights infringements for various others amounts to treating this person as a mere means. Utilitarians in particular could claim it is not. Utilitarians insist that everybody should count for one and nobody for more than one. This means that before ever coming to the conclusion that someone’s rights ought to be violated for the greater good, they take everybody’s interests into account and thereby treat each person as an end. This includes the person on whom the burden might get imposed. Imposing a burden on someone for the greater social good therefore does not imply that the person has been treated as a means only. He has received equal consideration as an end in the decision process, just as everybody else has.
Nozick himself hints at this utilitarian challenge by asking: “But why may not one violate persons for the greater social good?” He answers the question with an appeal to the separateness of persons:

Individually, we each sometimes choose to undergo some pain or sacrifice for a greater benefit or to avoid a greater harm: we go to the dentist to avoid worse suffering later; we do some unpleasant work for its results; some persons diet to improve their health or looks; some save money to support themselves when they are older. … Why not, similarly, hold that some persons have to bear some costs that benefit other persons more, for the sake of the overall social good? But there is no social entity with a good that undergoes some sacrifice for its own good. There are only individual people, different individual people, with their own individual lives. Using one of these people for the benefit of others, uses him and benefits the others. Nothing more. … Talk of an overall social good covers this up. (Intentionally?) … He does not get some overbalancing good from his sacrifice, and no one is entitled to force this upon him…

These thoughts are familiar from previous chapters. Now we have to investigate the role of this move in Nozick’s argument, as well as what would happen to Nozick’s further position if one were to apply a reductionist response to his argument.

---

140 Nozick, Anarchy, State and Utopia, p. 32.
141 Nozick, Anarchy, State, and Utopia, p. 32-33.
5.1.1 Taurek and the View that Separateness Speaks against Adding in All Cases

First, we have to ask how exactly these considerations about separateness are supposed to support deontological side-constraints and the idea that we ought not to kill in order to save several lives. In the passage, Nozick argues: “…there is no social entity with a good that undergoes some sacrifice for its own good.” This must mean there is no subject that would enjoy the aggregated benefit of rescuing the many by violating the rights of one, so that comparing the number of people affected makes no sense.

The passage corresponds to passages by John Taurek who also seeks to defend the deontological constraint at hand by appeal to the separateness of persons. In his very similar wording the alleged role of the separateness of persons is even more explicit. He writes: “This [the deontological thesis] reflects a refusal to take seriously in these situations any notion of the sum of two persons’ separate losses.”142 A page later he adds the following thought about the possibility of ignoring side constraints: “I do not see that I can thereby spare a single person any greater pain.”143 Taurek’s idea seems to be that there is no single entity or person who will suffer the disadvantages. This allegedly means that we ought not to look at the aggregated amount of suffering. Separateness is supposed to speak against aggregation. In this sense, Taurek’s interpretation of the appeal to separateness is a negative argument against views that support aggregation.

However, it is not clear that Nozick and other deontologists have to oppose adding in literally all cases. If we could either rescue one person or rescue five, in both cases without violating anybody’s rights, then most deontologists would presumably want to say that it is

---

142 Taurek, “Should the Numbers Count?” p. 308
143 Taurek, “Should the Numbers Count?” p. 309.
better to rescue the five. For example, if we could either rescue one person standing on a rock in the middle of a rising tide or five persons standing on another rock, most deontologists would say that we should rescue the five (or at least that it is permissible to save the five.) Opposing adding in all cases means that we have to be indifferent when confronted with this choice. Taurek’s use of the “social entity” as implying that we should not add up under any circumstances is a rather extreme position. He suggests tossing a coin in these cases to give everybody an equal chance of surviving. Most deontologists will want to say something more moderate. They will want to say that we should avoid adding only in cases in which rescuing the five would require violating someone’s rights, for example by requiring us to actively kill the one. If we have to decide whether to help one person or a group of five, and we do not have to violate a right in either case, we should help the five.

This raises our previous question again. How does the separateness of persons speak in favour of the more moderate deontological position? Obviously, the claim that the separateness of persons speaks against adding in all cases is not able to establish the connection with the moderate position, because this claim will lead to the extreme position held by Taurek. If Nozick wants to hold the more moderate deontological position he needs to argue that we should avoid adding only if this would require an impermissible use of another person, as in the case of killing one to rescue five.

The emphasis on the impermissible active killing of one person for the sake of others in this formulation of side constraints seems to direct our attention at actions in addition to separateness. Now we have to ask how a relationship between separateness and actions could get fleshed out. For one thing, we first have to specify that the supporting relationship would have to run from the separateness of persons to my actions, rather than to actions in general. Otherwise, we would be back to where we started. We would face the question of whether it
might not be irrational to reduce the total number of murderous actions if we can do so by committing one mortal action. So to give this alternative proposal any chance of success, the supporting relationship would have to run from separateness to a prohibition of my mortal actions.

In order to do this we might have to view the separateness of persons from a first-person perspective. The proposal could then run as follows. I have first-person access to the fact that I myself am a separate entity. Of course I can assume that everyone else sees themselves as separate as well, but this is a second step. What is most obvious about the separateness of persons from my point of view is that I myself am separate. Consequently, the constraint would say that I have reason not to transgress the separateness divide between myself and all the other agents out there, no matter how many others there are. We may stick to some general terminology and call this perspective agent-centred. If this is our perspective on separateness, then the moderate side-constraint view might indeed follow. Saving the larger number would be impermissible whenever it would require that I actively commit an illegitimate act against another agent.

This is an interesting result. It is interesting because it seems to be in conflict with the way in which Nozick’s view is generally categorized. The language he uses suggests that he holds a patient-centred, rather than an agent-centred deontological view. The patient-centred view demands that nobody should be treated in an impermissible way and that it is impermissible to treat persons as a mere means to one’s own ends. The agent-centred view, on the other hand, focuses on our capacity as agents and says our primary duty consists in not
violating any rights by never committing an impermissible act, or in not being a certain kind of agent.  

When trying to construct an argument in favour of moderate deontological side constraints, it turns out that the patient-centred view on its own does not yield the desired result. The patient-centred view alone only gives us reason to avoid the use of persons as mere means. Of course our killing one person in order to rescue five would require using him as a means for rescuing the others. But in a relevant sense the rights violations against the five will be exactly the same. Someone will use them as a mere means by killing them. Since the patient-centred view on its own only gives us reason to avoid the use of persons as mere means, it might well demand that we ought to avoid uses as mere means of the larger number.

The only way of avoiding this that we have seen so far would be to follow Taurek’s interpretation of the appeal to separateness as demanding a refusal to add in all cases, because there is no social entity that will enjoy the aggregated benefit of the outcome. But this will lead to the extreme position that we should avoid adding even in cases in which no rights violation on our part is involved. The further condition that is needed in order to arrive at the more popular moderate view is an additional focus on the agent-centred understanding of deontology.

The focus on the agent-centred understanding of deontology might well be supportable by the appeal to the separateness of persons. The separateness of persons implies that our relationship with our own self is different from our relationship to any separate person. If my special relationship with myself therefore directs the attention at my actions rather than anyone else’s, this might support the idea that from a first-person perspective my actions matter more than anyone else’s. Proponents of side-constraints could argue that it is a deep

---

and metaphysically fundamental fact that my actions are mine, authored in a fundamental sense by me and not by anyone else.

However, being mostly concerned with my own actions can be seen as ego-centric in a disturbing way. The reason is that it can lead to egoism if it is not supplemented with further convictions. If in addition to caring mostly about my own actions I also believe that the end of these actions should generally involve a benefit to myself, then we might be looking at an argument in favour of egoism rather than side constraints.

On the other hand, if in addition to caring mostly about my own actions I also give a smaller amount of weight to the patient-centred concern for other people’s interests, then the view can support side-constraints. A strong emphasis on the agent-centred view, in combination with a somewhat smaller amount of weight for the patient-centred view, can yield the desired result in cases where we have to decide whether to kill one person in order to save the five. If the fundamental metaphysical nature of agency gives the needed strong weight to the view that I ought to care primarily about my own actions, while also giving a smaller amount of weight to the patient-centred concern for others, then I ought not to kill, irrespective of whether this leads five other persons to commit murder. The reason is that my stronger concern not to commit an impermissible action myself will outweigh the concern of rescuing a larger number of patients. This is the view we need in order to arrive at moral side constraints.

Nonetheless, Nozick mostly talks as if he intends primarily to support the patient-centred view. He regularly insists that a person is “not to be used for any end except as he chooses” and argues that “distinct individuals … are not resources for others.”

The numerous passages in which he uses such language all seem to suggest a patient-centred

---

145 Nozick, _Anarchy, State, and Utopia_, pp. 31, 33.
perspective on separateness. They suggest what makes an action impermissible is that it would be wrong to be *subjected* to this action. Given this emphasis in Nozick’s text it is entirely implausible to assume that he intends *only* the agent-centred view. At the very most we could say that he holds the agent-centred view *in addition* to the otherwise emphasized patient-centred view. It is somewhat hard though to read his text as a supporting the agent-centred view *more* than the patient-centred one, even though this is what he should support.

### 5.1.2 Agent-Centred and Patient-Centred Perspectives on Separateness

We have already seen how the agent-centred view might follow from the separateness of persons. Since Nozick obviously holds the patient-centred view—either exclusively or in addition to an agent-centred view—we still need to ask whether the patient-centred view also follows from the separateness of persons and whether anything about the relationship between the two views follows from separateness. Nozick insists rather explicitly that the patient-centred view indeed follows from the separateness of persons. What makes the mere use of others wrong is that they are separate persons, each of them with their own lives to lead. As he argues in the passage quoted above, we sometimes sacrifice our own present good for some reason, but it seems wrong to make such a choice for others without their consent. The idea is that we do not have the moral authority to decide which burdens they should be willing to undergo for any particular personal or social end and it seems to be precisely the separateness of persons that makes this the case.

There is an interesting relationship between the patient-centred view and an agent-centred view that seems relevant here. The patient-centred view says that each person has separate experiences and that it is therefore wrong to treat a person as a mere means by ignoring whether they would consent to an action that affects them. This view seems to receive extra support from the idea that persons are also separate agents who wish to make
their own choices when it comes to their lives. In this way the separateness of agents actually complements and supports the patient-centred view.

This observation is related to one of our earlier observations about the agent-centred view. In a certain sense the agent-centred view also seems to receive additional support from the patient-centred one. When asking why it should be bad to commit a certain action or to treat someone in a particular way it helps to refer to patient-centred considerations. It is due to their capacity as recipients of experiences that persons find it disagreeable to be subjected to certain kinds of treatment.

It is also interesting to see that both the agent-centred and the patient-centred view seem to have a first-person and a generalized aspect. My first-person perspective on my own agency merely tells me that I myself seem to stand in a particularly intimate relationship with my own actions. Likewise, a first-person perspective on my own capacity as a recipient of experiences and treatment by others merely tells me that certain experiences are agreeable for me and others are not. Both of these first-person perspectives can lead to the conviction that I should favour myself. In other words, when considered from a first-person perspective, both views risk collapsing into egoism.

According to both the agent-centred and the patient-centred view only the further conviction that others have a first-person viewpoint onto their own agency or experiences as well can yield a moral view other than egoism. Only this further belief about others can lead to the conviction that we ought to respect the first-person agency or experiences of others, which is a central requirement of moral views other than egoism. In this sense, both the agent-centred and the patient-centred view can be subdivided into a first-person and a generalized view.
The only perspective that seems to support side constraints, however, is the first-person agent-centred perspective. This is the only perspective that tells us preventing rights violations committed by ourselves is more important than the autonomous experiences and agency of others. As soon as I place equal weight on the generalized fact that others want their own experiences and agency preserved as well, it becomes less plausible to say that I should only worry about violations that originate from me, and more plausible that I should worry about everybody’s agency and experiences. This recognition would have to direct my attention at the disagreeableness of interferences in general, not only at the ones that originate from me.

Even though putting a large amount of weight on the first-person agent-centred perspective is the only way in which we can generate support for side constraints, this perspective is always at risk of collapsing into other perspectives that lend less support to side constraints. The reason is that we need to combine it with a somewhat smaller amount of weight for either of the generalized views in order to avoid egoism. Otherwise, we would have no reason for believing that we ought to avoid certain kinds of actions against others in the first place. However, understanding that it is problematic to commit certain kinds of actions against others, because they are separate agents or separate experiencers like ourselves, implies appreciating the fact that a first-person agent-centred perspective is not the only one there is. The conviction that I ought not to use others in certain ways could not follow if I did not consider a generalized perspective of either kind. Interestingly, both kinds of generalized perspectives give me reason to worry about everybody’s agency or

146 Both deontological as well as utilitarian authors have made similar observations. Compare for example Nagel’s argument on why the “solipsism of the egoist” is problematic in Nagel, The Possibility of Altruism, part III. Likewise, Sidgwick argues a recognition that one’s own good is part of a larger good would have to compel the egoist to adopt a less self-centred perspective. Sidgwick, The Methods of Ethics, pp. 420-21.
experiences, not just about actions that originate from me. This means they might well give me reason to rescue the larger number of agents rather than avoiding the killing of the one.

5.1.3 Can Separateness Support Side Constraints that Avoid Taurek’s View?

We have now seen that Nozick’s text passages gives us reason to think he holds a (generalized) patient-centred view deriving from the separateness of victims, while his argument also requires him to give predominant weight to a first-person agent-centred view deriving from the separateness of agents. We may now ask how much weight he can actually give to his patient-centred view in addition to the first-person agent-centred view without endangering his own argument in support of side constraints.

There are two possible ways in which Nozick can proceed here, both of which are somewhat unsatisfying. First, one can argue that it is possible to give an equal amount of weight to all kinds of separateness without endangering the argument in favour of side constraints, if we add Taurek’s position. We may give a serious amount of weight to the first-person separateness of agency. This supports avoiding the one killing. At the same time, we may give a serious amount of weight to the other aspects of separateness, which may speak in favour of killing the one to rescue the many. Now we also add Taurek’s thought to this: There is no social entity that will enjoy an aggregated outcome, which means even if we are concerned about all victims, we have no reason to benefit the greater number. Taken together, this means we ought to avoid killing the one in order to rescue the many.

In other words, we might not actually have to be more concerned with the first-person separateness of agents than with the other aspects of separateness if we want to achieve the desired result. Instead, we can be equally concerned with all kinds of separateness, and it can be Taurek’s objection against adding that tips the scale in favour of avoiding the one killing.
We just need to combine an equal concern for all kinds of separateness with Taurek’s concern about adding in order to achieve the desired result.

This combination of views is a serious option. However, we need to be aware that this reading of Nozick’s view will also lead to Taurek’s extreme conclusions. If the objection against adding applies in the case at hand, it also has to apply in cases in which nobody’s rights are at stake. When deciding whether to rescue five people by doing something perfectly permissible, or rescuing one person by doing something perfectly permissible, we would have to toss a coin. This will seem counter-intuitive even to most deontologists.

Most deontologists will want to hold that we only have to object to adding when somebody’s rights are at stake, as in the case in which rescuing the five would require a killing. It remains entirely unclear why the separateness of persons should speak in favour of the objection against adding in some cases, but not in others. Why exactly should the objection against adding only apply whenever a rights violation would be involved in maximizing the outcome?

The only answer available seems to be that there is something particularly fundamental about first-person agency that makes it the case that we primarily have to avoid any rights violations committed by ourselves. We are free (or even ought to) add whenever this would not require a rights violation. According to this view, “keeping our own moral house in order” is so important that it outweighs the aggregated advantage of rescuing a larger number of people. This combination of views is also a serious option.

However, if this is our view we also need to be aware of an important disadvantage. Arguing that “keeping our own moral house in order” is more important than the otherwise also important aggregation of benefits amounts to the claim that the first-person separateness of agency is more important than the other aspects of separateness—a weight distribution for
which we have no argument. Why should the first-person separateness of agents be more important than the separateness of victims? Nozick’s argument is supposed to derive support from a general understanding of the separateness of persons. But on this reading it seems as though one kind of separateness is given more weight than a parallel and equally plausible kind of separateness that would actually support a different moral decision if it were given more weight. This seems arbitrary if we cannot find an argument for it.

My worry is that this kind of asymmetry is only plausible if we give the needed extra bit of weight to the agent-centred view for reasons that have nothing to do with the separateness of persons. People often believe that doing something is morally more important than letting something happen without tying their view to the separateness of persons. If we are tempted by this thesis irrespective of its relationship to the separateness of persons, then we might be able to give the needed extra bit of weight to the agency view. But this extra little bit of support does not seem to derive support from a holistic understanding of the separateness of persons.

5.2 How Separateness Might Support Contractually Derived Basic Rights and Liberties

According to Rawls the separateness of persons supports the idea of a contractual method to derive the moral principles that should structure a society. He also believes that a proper contractual method will yield his two principles of justice at least the first of which is deontological. He seems to believe there is a positive argument from the separateness of persons for contractualism and consequently for his principle of equal basic rights and liberties.\(^\text{147}\) In laying out this positive argument I will be focusing on *A Theory of Justice*. His

later views about the relationship between personal identity and normative philosophy are one of the subjects of the last chapter.

In order to establish a connection between the separateness of persons and a contractualist decision procedure, he argues that the competing method of using the standpoint of an impartial spectator is flawed. This method is used by classical utilitarianism, he argues. 148 Utilitarian benevolence does not offer an appropriate decision principle when the interests of separate persons are in conflict and “[b]enevolence is at sea as long as its many loves are in opposition in the persons of its many objects.” 149 In this strategy of attacking a competing view, his argument in favour of contractualism is at least partly a negative argument.

One straightforward response to the negative argument in favour of contractualism seems to be that benevolence might offer a decision procedure after all: for example the decision procedure of classical utilitarianism. The impartial benevolent spectator could simply decide to maximize the aggregated outcome. If two or more people have conflicting interests, the spectator could always decide in favour of that outcome which will maximize overall subjective utility. In this way benevolence would have maximally reached its goal.

Rawls would object that in such a procedure “the principle of choice for one man is applied to society,” which allegedly amounts to “conflating all persons into one.” 150 As discussed in chapter two, utilitarians can deny this. Parfit for example does not argue that the principle of choice for individuals has to be used for societies on the ground that society is a

mass person. He does not believe society is a mass person. Instead, he acknowledges the disintegration within lives in addition to the separateness between lives.

Utilitarians will also insist that their principle of choice for society is by no means derived from the principle of choice for individuals. Instead they believe their principle of choice for societies is a self-standing independent principle that is on a par with the principle for individuals. As McKerlie points out, the two principles are different in the sense that they have a different scope. One of them maximizes within lives, the other in societies. Utilitarians do not tend to argue that individuals must maximize welfare within their lives. Parfit’s view gives room for using the principle of equality within lives. McKerlie also adds: “Rawls must convince us that the difference in scope is not itself enough to accommodate the undeniable difference between an individual and a society, but nothing in the argument performs that task.”

I have said more about the negative argument against utilitarianism in a previous section. Here we should investigate whether anything resembling a positive argument in favour of contractualism can be derived from Rawls’s appeal to the separateness of persons.

5.2.1 Contractualism and Respect for Hypothetical Separate Persons

Rawls seems to imply that there is a positive argument between the separateness of persons and contractualism. In section 30 of *A Theory of Justice* he argues:

… a love of mankind that wishes to preserve the distinction of persons, to recognize the separateness of life and experience…is guided by what individuals themselves would consent to…

Rawls’s concern is a strong regard for the notion of individual consent. A decision procedure from a third-personal standpoint is not reliably able to guarantee this, he believes. However, it is important to remember that a solution all conflicting parties would consent to might not always be reachable. Contractual considerations can seek to establish at the most a *hypothetical* agreement about the *general* principles by appeal to which the decision should be made. To ensure a solution will be generated, Rawls stipulates that the contracting parties are similar to each other in the sense that they are rational and willing to reach a solution. He also abstracts from various other potential differences between the parties that could prevent a successful settlement. In his well-known description of the original position, he thus describes the contracting parties as follows:

…no one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like. Nor, again, does anyone know his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism. More than this, I assume that the parties do not know the particular circumstances of their own society. That is, they do not know its economic or political situation, or the level of civilization and culture it has been able to achieve. The persons in the original position have no information as to which generation they belong.\(^{153}\)

Two pages later he adds:

Therefore, we can view the agreement in the original position from the standpoint of one person selected at random. If anyone after due reflection prefers a conception of justice to another, then they all do, and a unanimous agreement can be reached.  

I will not rehearse familiar criticisms of these features of the original position. Critics from the libertarian right as well as from the communitarian left have objected to this understanding of the person in the political decision process. Rawls’s original concern was that a third-personal decision procedure would not be able to yield principles that would take into account what the contracting parties themselves would consent to. As it turns out, however, the contractualist method itself is only able to ensure this to a limited extent. If we were to ask what actual contracting persons would consent to, then we would often not reach a decision at all. This is why Rawls needs to construct a hypothetical contract situation between persons that are stripped of personal differences that could prevent a settlement. In this sense, the contractualist method only seems to respect the separateness of hypothetical rational persons rather than the separateness of actual conflicting persons. If the idea is that every actual person should consent to particular ways of settling things, then the appeal to separateness is inherently normative. I will say more about this in the next chapter.

5.2.2 How a Benevolent Spectator Might Also Respect Separateness
We still have to ask what exactly it is about the notion of consent in Rawls’s contractualism that cannot be accomplished by a benevolent spectator. Rawls seems to believe a benevolent spectator cannot reach a decision when his many loves are in conflict without disregarding the

---

154 Rawls, A Theory of Justice, rev. ed., p. 120.
155 Dworkin, “Justice and Rights.”
separateness of persons. Our earlier concern was that benevolence might be able to settle things after all. For example, a utilitarian framework would seek to maximize subjective welfare without placing special value on consent. Likewise, alternative forms of consequentialism would seek to promote the good rather giving exclusive or primary importance to consent.

Presumably the contractualist concern with these principles will be the potential use of a notion of the good or a principle that the parties themselves might disapprove of. Standing in a conflict with another person and seeing this conflict settled by appeal to a decision procedure that one disapproves of is certainly problematic. It is more satisfying to experience a settlement that is reached via a decision principle that one is willing to accept, even if the settlement as such is disadvantageous for oneself.

However, a benevolent spectator would presumably take this into account. Wanting the best for the parties involved implies taking into account that a settlement achieved via a decision procedure that one or more of the parties cannot accept will be significantly less benevolent than a settlement achieved via a decision procedure that both parties can approve of. While benevolence can in principle appeal to conceptions of the good or to principles that the parties disapprove of, it is arguably not an act of benevolence to impose these standards from above like a paternalist dictator. Benevolent parents trying to settle a fight among their children will usually appeal to reasons that even those children who will not get their way should understand.

Rawlsians might find such an understanding of benevolence unproblematic, but might also insist that a benevolent spectator who settles conflicts by appeal to procedures the parties can approve of is implicitly using a contractualist method. Supporters of the benevolent spectator, however, may respond to this in a similar way. They may insist that asking the
contractualist question of what the parties would consent to is just a different way of wondering what, objectively speaking, the best principle or the best conception of the good for the settlement of the conflict consists in. Benevolent deliberation about the best moral positions may well be sufficient to generate the realization that a paternalist dictatorship is not an appropriate moral procedure.

If benevolence implies taking the convictions of the parties into account, we may have reached a stalemate between the two approaches. To the charge that such a form of benevolence is a hidden form of contractualism, supporters of benevolence can respond that contractualism is just an elaborate way of asking what the right benevolent principles are. If we have the choice between contractualism and a benevolent spectator who takes the convictions of the parties about his decision procedure into account, then what exactly will contractualism add to the debate?

Persons in the original position consent to a principle if that principle accords with what they are willing to choose for their future society. So what is meant is not merely a passive form of acceptance of a future principle, but a deliberative, active decision in favour of particular principles and against others. This gives rise to the impression that contractualism might place special weight on the separateness of agents. People’s choice behind the veil of ignorance is an action rather than a passive experience of particular principles as appropriate. So if the choice is between contractualism and a benevolent spectator who takes the convictions of the parties about an appropriate decision procedure into account, then contractualism might add a more or less symbolic appreciation of people’s capacity as free and autonomous agents.

If one were to place more importance on people’s capacity as patients, the motivation for a contractualist setup might not arise to the same extent. The choice of a contractual
method as such seems to imply that people’s capacity for making free choices outweighs all their other properties and capacities and warrants a special additional device in our theory. It is possible to reject this weight distribution between the separateness of persons as agents on the one hand and the separateness of persons as the recipients of experiences and meaning on the other. For example, utilitarians can insist on the human capacities for welfare and suffering as the most fundamental aspects of a theory. They strongly acknowledge the value of autonomous individual choices, but they also insist that the value of these choices lies in their contribution to human happiness and welfare. The capacity for welfare and suffering implies a great regard for people’s free and autonomous choices, but it does not necessarily add enough appeal to the idea in order to construct a contractualist framework for our theory. In fact, some of the most outspoken opponents of paternalism and supporters of individual choice have been utilitarians who nonetheless abstained from formulating contractualist frameworks.

A holistic understanding of the separateness of persons that does not give disproportional weight to people’s capacity as autonomous decision makers and agents does not necessarily lend sufficient support to contractualism in and of itself. In order to yield sufficient support for contractualism we would need an argument to the effect that the separateness of agents is more important than the separateness of patients or other aspects of the nature of persons. It would have to explain why the primary value of autonomous agency might not consist in its conduciveness to promoting other values, such as experiences of meaning and respect. Even if such an argument can be given, though, we may still wonder whether a more holistic understanding of the separateness of persons might not lead to other results.
5.2.3 Equal Rights and Liberties: Separateness as a Coarse Tool

In the original position Rawls’s separate persons decide in favour of a principle of equal basic rights and liberties. After looking into the relationship between separateness and the motivation for contractualism, we should ask whether there is a special relationship between separateness and the choice of equal basic rights and liberties. Rawls not only seems to believe that separateness might speak in favour of contractualism, but also that the contractualist method, if engaged in by separate persons, will lead to a principle of equal basic rights and liberties.

The distinguishing feature of Rawls’s principle of basic equal rights and liberties consists in the fact that these rights and liberties may not be traded for distributive gains. For example, he would object to giving any small amount of extra rights and liberties to those who manage to create improvements of everybody’s opportunities as well as improvements of the situation of the worst off. Rawls believes the separateness of persons speaks in favour of contractually derived equal basic rights and liberties that cannot be traded against such other advantages.

As already pointed out in the chapter dealing with consequentialism, the separateness of persons might support an initial assumption of distributive equality and an initial assumption of priority for the worse off. It also turned out, however, that such an initial assumption of equality might not tell us much about the specific way in which we should respond to it. As argued, it is unclear how an argument in favour of equality of opportunity, rather than some competing form of equality, could depend directly on a holistic understanding of the separateness of persons. Moreover, in a way even utilitarians seem to account for an initial equality assumption when insisting that everybody should count for one and nobody for more than one. In addition, the diminishing marginal utility functions of most
instrumental goods might effectively tend to promote equal distributions rather than unequal ones even within a utilitarian framework.

Now we are faced with the question how the separateness of persons might speak in favour of equal rights and liberties that cannot be traded against other egalitarian principles that the separateness of persons allegedly also supports. Clearly, there can be conflicts between equal basic rights and liberties on the one hand and distributive equality and priority assumptions on the other hand. If the separateness of persons speaks in favour of equal basic rights and liberties in addition to distributive equality, then we are faced with the priority or ordering question in cases of conflict.

Can appeals to separateness settle such ordering conflicts? The fact that separateness apparently supports both rights-related and distribution-related equality inclinations may make us skeptical. It looks as though we are missing an argument from a holistic understanding of separateness to an ordering principle. If the separateness of persons cannot tell us in which ways our initial assumption of equality should enter our moral theory, then it might also not be a sharp enough tool to help us decide between various moral theories that give weight to equality in some way. Without further refinements a general appeal to the separateness of persons might be a rather coarse tool when it comes to equality issues. It seems to be able to give rise to an initial assumption in favour of equality without telling us much about how to respond to conflicts between the various ways in which such equality considerations can get fleshed out.

What might complicate matters further are causal effects from potentially permissible economic inequalities to the worth of people’s equal liberties. Daniels, “Equal Liberty and Unequal Worth of Liberty.”
6 What Can Appeals to Personal Identity Establish in Normative Ethics?
The previous chapters have discussed a number of controversial arguments from personal identity to particular normative conclusions. Overall, it has turned out that the relationship between personal identity theory and ethics is often more fine-grained than philosophers have initially presumed. The support that appeals to the nature of persons can lend to particular moral views seems to depend to a considerable degree on the chosen set of criteria for identity as well as on the amount of weight one is willing to assign to the various possible criteria.

This fine-grained structure of arguments from personal identity calls into question whether entire normative theories like utilitarianism or Kantianism can ever receive conclusive support from appeals to personal identity without further specifications about the specific amount of weight one is willing to assign to an entire array of criteria. As a consequence, there have been a number of attempts to assign an alternative and more modest role to personal identity appeals. Two such attempts have been presented by the later Rawls.

This later view on the subject can be subdivided into at least two central ideas that represent important ways in which one can be skeptical about arguments from personal identity theory to ethics: the independence thesis and the plasticity thesis. The independence thesis says moral theory is actually independent of personal identity theory. The plasticity thesis says the relevant dependence relationship actually runs the other way around: from moral theory to personal identity criteria.

6.1 The Independence Thesis
6.1.1 Rawls’s Formulation of the Independence Thesis

To figure out whether moral theories might actually be entirely independent of personal identity theory, it is important to take a more exact look at the argument to that effect. This is how Rawls puts the thesis in “The Independence of Moral Theory”:
…the conclusions of the philosophy of mind regarding the question of personal identity do not provide grounds for accepting one of the leading moral conceptions rather than another. Whatever these conclusions are, intuitionism and utilitarianism, perfectionist and Kantian views, can each use a criterion of identity that accords with them. Thus while the philosophy of mind may establish conditions that any correct criterion must satisfy, none of the traditional doctrines are affected by these constraints, at least not so long as these doctrines are applied under the normal conditions of human life.\footnote{Rawls, “The Independence of Moral Theory,” p. 15.}

Then he adds:

…there may be variations among these criteria, different views counting certain features of the person as more important than others. At the same time, however, all the criteria satisfy the conclusions of the philosophy of mind…\footnote{Rawls, “The Independence of Moral Theory,” p. 15.}

6.1.2 Two Readings of the Independence Thesis

In interpreting these passages it turns out that there are at least two possible ways of specifying where exactly the ground for the independence lies. The source for this ambiguity seems to be the expression “the conclusions of the philosophy of mind”. This expression sounds as if there were a consensus in philosophy of mind about the central constraints, which might well be wrong. If there is relevant controversy in philosophy of mind, and there are
several leading theories about the relevant constraints, then it turns out that Rawls’s characterization of the independence thesis allows the following two readings:

(A) Each moral theory recognizes a criterion that is compatible with all theories of personal identity that the philosophy of mind has established.

(B) Each moral theory recognizes a criterion that is compatible with the constraints of at least one theory of personal identity that the philosophy of mind has established.

This ambiguity only arises if there really are several theories of personal identity that cannot reach a consensus about the constraints. In other words, if we do not want to conclude that Rawls was ambiguous, then it had better be the case that there is a relevant consensus about the constraints among the theories of personal identity. But is there consensus?

Rawls indicates there is such a consensus, “as long as these doctrines are applied under the normal conditions of human life.” This expression actually seems to suggest that Rawls—quite naturally—focuses on those controversies that are discussed at around the time of his conception of the independence thesis rather than on accounts that were introduced later. At that time there was actually a lively discussion in personal identity theory that had notoriously little bearing on real life moral questions. The conflicting positions lead to the same judgments when “applied under the normal conditions of human life.” This controversy regarded the question whether the physical persistence of the brain should be a necessary criterion of personal identity. Neo-Lockeans wanted to remain agnostic about this, while the even more scientifically minded wanted to include physical persistence among the necessary criteria of personal identity. Despite being fundamentally at odds about this question, both
proponents and opponents of the physical criterion agreed that psychological continuity should be a necessary criterion, which is why they produced the same answers for real life cases.

The arguments on either side lead to some of the most hypothetical thought experiments in the history of philosophy. Only with regard to these hypothetical cases do the accounts produce different judgments about the persistence of persons. One of them is the thought experiment of body transfer or teletransportation. In this thought experiment, we imagine that the physical constitution of a person’s body and brain get recorded by a scanner. Then the person’s body gets destroyed. At the same time, a replicator on a different planet produces a new body according to the recording. After that being wakes up, it has all the recollections of the original person on earth. Are we dealing with the identical person? A psychological continuity account can maintain we do. An account requiring continuity of brain or body will have to conclude the person on the other planet is a different being.

A couple of paragraphs after his characterization of the independence thesis, Rawls considers the extent of the controversy in personal identity theory and addresses exactly this debate. He correctly concludes:

…it is unclear whether bodily continuity is a necessary part of personal identity. But if we leave aside such hypothetical cases as fission and bodily transfer, cases that we use to explore criteria of identity, then the main reasons for thinking that these conditions will give the wrong results are theological. … But even if within a secular framework
bodily continuity should prove unnecessary, my surmise is that this would still not render one moral conception more reasonable than another.\textsuperscript{159}

Rawls is right in pointing out that the debate about physical continuity created different judgments only for cases like fission and bodily transfer. In regard to the independence thesis this means the controversy had no obvious bearing on moral theory.

However, since the 1990s some philosophers have introduced new accounts of personal identity that give quite different answers to the question under which conditions a person remains identical. These newer accounts do have a bearing on real life cases and therefore potentially also on our choice of moral theory. One of the most notable suggestions of this sort is Eric Olson’s animalism, according to which the necessary persistence condition of persons is the persistence of their animal organism. The animal organism in turn remains identical as long as there is continuous persistence of that part of the brain that organizes its life sustaining functions, which Olson takes to be the brain stem.

The most obvious example of a case in which this account makes a different judgment than a psychological continuity account is the one of a human being in an irreversible vegetative state. In this case, the animal organism has remained preserved, but there is no psychological life anymore. Olson would therefore say that the person has remained identical, while a psychological continuity account would have to say that the person has ceased to exist.

According to animalism, the \textit{psychological} life of the person is merely a \textit{property} that the animal organism might or might not have at various points in time. Animalism treats it as a so-called phase sortal rather than a substance sortal. Whether the organism has lost this

\textsuperscript{159} Rawls, “The Independence of Moral Theory,” p.16.
property does nothing to determine whether the person has remained identical. As a consequence, even a being in an irreversible vegetative state is still the same person as before.

There are also more nuanced approaches. For example, Jeff McMahan suggests the necessary conditions under which persons remain identical are the continuous persistence of those parts of the brain that are responsible for uniquely personal mental states, as well as some basic psychological functioning of those parts. This allows McMahan to maintain that a late stage Alzheimer patient who displays no psychological continuity anymore but still possesses at least momentary consciousness is still the identical person—a view that might not be open to a psychological continuity theorist. At the same time, he can say that bodies in persistent vegetative states are not persons anymore, because they do not display any basic psychological functioning anymore.

As McMahan’s account shows especially well, one does not have to adopt diachronic psychological content as a necessary criterion of personal identity—not even if one takes a psychological component to be necessary. One can reduce the necessary psychological component to more basic notions such as minimal psychological functioning. This minimal functioning could be as basic as a mere capacity for momentary sense impressions. However, a requirement for momentary consciousness alone does not give us any \textit{diachronic} persistence conditions of a being. Since this is precisely what an account of personal identity is supposed to give us, though, it is necessary to combine it with a diachronic persistence condition of something physical.\textsuperscript{160} In this way, we can get a diachronic account even without requiring any diachronic psychological content.

\textsuperscript{160} Parfit anticipates a somewhat related possibility in his \textit{post-Reasons and Persons} article “Experiences, Subjects, and Conceptual Schemes,” p. 218. There, he briefly suggests combining a momentary consciousness requirement with animalism.
In any case, the existence of these newer accounts, which do make a difference for real life cases, supports the need for looking at two readings of the independence thesis. It is quite possible now for a moral theory to endorse a criterion the real life consequences of which are not automatically compatible with the consequences of all other criteria that personal identity theory has established.

There is a further striking observation to be made. Given the newer accounts, the relevant discipline in philosophy when it comes to personal identity theories is not just philosophy of mind any more, as Rawls still assumes. We also have to look at accounts of the organism that the philosophy of biology generates. In addition, we might have to look at what medicine establishes about the functions of the various parts of the brain. I will therefore suggest a more cautious wording of the two readings of the independence thesis that will not claim that we need to look only at what the philosophy of mind has established. The more cautious formulations would thus be these:

(A') Each moral theory recognizes a criterion that is compatible with all theories of personal identity.

(B') Each moral theory recognizes a criterion that is compatible with the constraints of at least one theory of personal identity.

To verify one of these readings, we should now take a more detailed look at the implications of the competing kinds of candidates for the necessary persistence conditions of persons.
6.1.3 Independence and Purely Physical Criteria

The most radically revisionary suggestion is probably the one made by Olson, according to which the physical persistence of the brain stem is the only necessary condition for the persistence of a person. It has been objected that such a view is not really an account of personal identity, but rather one of organism identity. Olson is unmoved by these worries, though. As the subtitle of his book suggests he takes his view to be an “account of personal identity without psychology.”

What is interesting about this criterion is that it might actually have no implications at all for normative ethics. Usually there needs to be some sort of psychological life in order for a moral theory to be applicable at all. Utilitarianism, for example, is interested in maximizing subjective welfare; deontological theories want to establish rightful relations between persons. An unconscious organism is not able to be the subject of either concern.

One could still ask whether there might nonetheless be something like an objective wellbeing of a potentially unconscious organism. An organism might be doing well when it is functioning in a way that its constitution seems to suggest as the proper way. In other words, it might be doing well whenever it functions in accordance with its telos.

However, which normative theory or which overarching moral principle would this support? Teleology in general? It could surely be the basis of a teleological theory, but only for a teleological theory of the organism. A teleological theory of the organism alone does not give us answers to many of the questions that we usually want answered by a moral theory. The telos of an organism does not tell us what the principle guiding our actions as conscious beings should be. At the most it might tell us that we should not interfere with and maybe even promote the proper functioning of human organisms. This could have bioethical
consequences. It could also be the foundation of a form of environmental ethics that is concerned with organisms. It could probably not be the basis of a full moral theory though.

Arguably, this criterion could also lend some initial support to equality. As soon as we can distinguish between different human entities, an urge toward equality might arise. On the other hand, one may wonder whether there should really be equality between unconscious entities. In any case, equality among all human organisms, including those in vegetative states, hardly amounts to a general moral principle that can have something relevant to say about a good number of the further questions we want answered. An initial assumption in favour of equality does not tell us what kind of moral equality view to support, nor does the organism criterion provide a basis for issues such as responsibility ascriptions.

In other words, if the physical persistence of the brain stem is the one and only necessary criterion of personal identity, then personal identity might not actually support a full moral theory at all. There would literally be close to no relationship between our necessary personal identity criterion and moral theory in the sense of a general moral principle guiding our actions as conscious beings. This finding, taken on its own, actually seems to support independence between the personal identity criterion at hand and a full moral theory.

What exactly does this mean so far for the two possible readings of Rawls’s formulation of the independence thesis? Version (A') says that each moral theory recognizes a criterion that is compatible with all personal identity theories. Version (B') said each moral theory recognizes a criterion that is compatible with at least one personal identity theory. So far, we have found out that there is one personal identity theory that seems to support no full moral theory at all. This finding is compatible with both readings of the independence thesis as long as we do not accept a teleological account of organism functioning or an initial bias in
favour of equality as a full moral theory. However, before basing too much on this finding, we should look at another possible kind of criterion.

6.1.4 Independence and Criteria Requiring only Minimal Psychological Content

The next criterion we should look at is physical continuity of particular parts of the brain with at least minimal psychological functioning. This is a compound criterion. What such a compound criterion adds to a physical criterion is the requirement of an at least minimal amount of consciousness. The capacity for an at least minimal amount of consciousness seems to allow for ascriptions of basic subjective wellbeing. This means we have a relation to one of the central components of a utilitarian theory in this criterion. Utilitarianism seeks to maximize subjective wellbeing. (Of course we also have a basis for a bias in favour of equality again.)

The sources of subjective wellbeing that could occur in a minimally conscious life would be rather impoverished, though. In normal conscious lives we can derive pleasure from holistic or long-term goods, such as the completion of a book or the fulfillment of a long-term plan. A criterion of personal identity requiring only minimal consciousness could not support a demand for maximization of holistic or long-term goods as important sources of pleasure. It would also not be able to take account of the potentially over-proportional badness of prolonged or anticipated suffering.

Nevertheless, one could argue these are only instrumental problems from the perspective of wellbeing. There might well be things the experience of which requires more extensive consciousness, and which give us a particularly large gain in wellbeing. But these are merely considerations about particularly efficient sources of wellbeing, not about wellbeing as such. So a criterion that requires merely minimal psychological functioning might actually be enough to support the central element of utilitarianism after all.
On the other hand, this criterion might also have a relation to an important ingredient of normative egoism. As soon as there is minimal consciousness, there might be self-bias. Since the kind of consciousness required by this criterion is very narrow, the kind of egoism it could support would also be of rather narrow character. For example, it would seem to support some form of egoism of the moment in case the consciousness is only momentary.

One may object that egoism of the moment is not a genuine moral theory. For one thing, egoism of the moment seems to be a rather incomprehensible and incoherent dogma. How could I possibly do whatever I want right now? By the time I would have started working toward the satisfaction of my desires, my relevant momentary consciousness would already have faded. So it is probably possible to rule out the idea that egoism of the moment is a relevant competitor to utilitarianism when it comes to the question which moral theory the criterion at hand would support.

Despite this, the fact that a capacity for minimal consciousness might also give rise to an (almost incoherently) narrow form of egoism also points out an important limit of the apparent support for utilitarian welfare. What would receive support from this criterion is the welfare part of utilitarianism, not the part that requires equal concern for all beings and the maximization of the general welfare. As argued in the chapter on consequentialism, a mere capacity for momentary consciousness cannot support the agent-neutrality of utilitarianism. A demand for agent-neutrality requires appeals to additional considerations.

One more thing needs to be added here. The criterion at hand would obviously not be able to lend support to moral theories making judgments about questions that can only arise for beings with more extensive consciousness. For example, it would not be able to support judgments about the conditions under which someone is morally blameworthy for a past action. Blame for past actions applies only to subjects with diachronic consciousness. A being
experiencing merely momentary sense impressions cannot be blamed for a past action. It fact, it cannot be said to have acted in the first place.

It is not the case, though, that a moral theory which assigns responsibility and blame for past actions would be literally incompatible with a criterion requiring only minimal consciousness. Even if the essence of a person consists in physical brain continuity with only minimal psychological functioning, we can still assign blame for past actions. However, in this case we have to base our judgments on more highly developed, but merely temporary properties of a person rather than on the one necessary property for the person’s identity. In other words, we would be basing our judgments on a mere phase sortal rather than a substance sortal which tells us that the necessary condition of personal identity is something else.

In this sense, there would actually be independence between our personal identity criterion at hand and the moral theory that requires us to assign blame for past actions. What would matter for this moral theory would not actually be a necessary property of persons without which we would literally cease to exist. Instead, the view would be based on a merely temporary property of persons that persons may lack without ceasing to be persons.

What does this mean for the independence thesis? Obviously, some moral theories want to assign blame and responsibility for past actions to persons. If we want to adopt such a moral theory, then there seems to be independence between our moral theory and a necessary criterion of personal identity that requires merely minimal consciousness. At the same time, though, the criterion does seem to lend some support to a utilitarian welfare theory.

This means reading (A’) of the independence thesis might well be discredited. It is not the case that all moral theories recognize all criteria that have been established by all theories of personal identity. A moral theory that wants to assign blame and responsibility cannot be content with a criterion that does not require complex and diachronic consciousness. Beings
that display no such consciousness are not the proper subjects of a principle making these kinds of judgments.

On the other hand, Rawls merely speaks of moral theories as “recognizing” one or the other criterion. A theory wanting to assign blame and responsibility might well “recognize” the importance of momentary consciousness, even if it requires diachronic consciousness in addition. There could not be any diachronic consciousness in the absence of momentary consciousness. In other words, the capacity for momentary consciousness might itself be necessary for diachronic consciousness, so that anyone who takes diachronic consciousness to be important would also have to recognize the importance of more basic forms of consciousness. This would mean that (A’) has not been discredited after all.

There is still a further consideration though. It seems as if a utilitarian welfare theory is able to do without a diachronic consciousness requirement. Arguably, it does not have to “recognize” a diachronic consciousness requirement. Moral theories giving intrinsic importance to blame and responsibility ascriptions have to recognize additional properties as important. They might be able to recognize the importance of minimal consciousness, but could not take minimal consciousness to be a sufficient criterion of personal identity.

This brings us back to our skepticism about (A’). There are moral theories that cannot recognize all criteria that personal identity theory has established. In particular, some moral theories cannot recognize a criterion requiring only minimal consciousness as a sufficient criterion of personal identity. So (A’) has been discredited. Only the claim that Rawls’s talk of “recognizing” was merely geared at necessary criteria, and not at necessary and sufficient criteria, could rescue (A’). Unfortunately there are no explicit text passages to this effect.

A general understanding of (A’) has to be wrong, though, on the ground that it works only for necessary criteria. It is not the case that there is independence between sufficient
criteria and moral theory. For example, if we choose continuity of brain with minimal momentary psychological functioning as a necessary and sufficient criterion, we have support for the welfare part of utilitarianism, but not for any theory that wants to assign blame and responsibility. This is a dependency relationship between a personal identity criterion and ethics.

6.1.5 Independence and Criteria Requiring Diachronic Psychological Content

A further modified version of (A') should still be addressed. It might be the case that all moral theories recognize all personal identity criteria as necessary and sufficient that require diachronic psychological content. Diachronic mental properties could be continuity and/or connectedness of first-person experiences or self-concern. Continuity and/or connectedness of beliefs, desires or intentions are also possible diachronic criteria. I know of no argument according to which a particular complete moral theory, apart from a teleological theory about organism functioning or the like, does not recognize one or more of these possible criteria as important at all. This means the following modified version of (A) could still be true:

\[(A'\prime\prime)\] Each moral theory recognizes a criterion that is compatible with all theories of personal identity requiring diachronic mental content.

Of course (A'\prime\prime) is a lot less interesting than (A'). As it turned out earlier, there can be criteria of personal identity that require no diachronic mental content or even no mental content at all. Moreover, organism criteria might lend support to teleological considerations or to a bias in favour of equality. Criteria that require only minimal consciousness seem to lend some support to utilitarian welfare considerations. So (A'\prime\prime) is simply not a very complete description
of the relationship between personal identity and ethics. It ignores some possible identity criteria.

(B') would have a stronger claim to being a complete description of the relationship between personal identity and ethics. According to (B') every moral theory seems to recognize at least one possible criterion of personal identity, so that no moral theory can get ruled out on the ground that it ignores literally every single criterion that has been established in personal identity theory. This means we now have to ask what to think of (B').

If (B') is true, then we can say that a general appeal to personal identity theory establishes nothing for moral theory. The reason is that such a general appeal does not tell us which criterion we are appealing to. Only once we have decided which (sufficient) criterion we find the most important or which relative weight we want to assign to an entire set of criteria will there be a dependence relationship with moral theory. Once we have decided which personal identity theory we take to be true and which criteria we subsequently take to be the most central and necessary elements of personal existence, there will be an impact on which moral theory we should support. Any appeal to personal identity in moral theory will be conditional on taking particular criteria to be the most relevant ones. Of course there should be good arguments for these criteria. After all, there are always other possible criteria that will support other moral conclusions.

If (B') is true, then there will not be complete independence between personal identity theories and moral theory. There will merely be independence between general, unspecified appeals to personal identity theory and moral theory. Once we have settled on a particular criterion of personal identity, though, our conviction might very well have an impact on moral theory. Only (A') would have guaranteed that it is irrelevant which criterion we settle on. It would have done so on the ground that all moral theories would have gone well with all
criteria. Nonetheless, it is very interesting and revealing that a general and unspecified appeal to personal identity establishes nothing for moral theory. In this sense there is independence between general personal identity appeals and ethics.

There is one way though in which complete independence even on a more specified level can still be true even if merely (B') can get established. Complete independence can still be true if it turns out that the arguments for one’s favourite criterion turn out to be appeals to the very moral theory that the criterion is supposed to support. This leads our discussion over to the plasticity thesis.

6.2 The Plasticity Thesis

6.2.1 Rawls’s Formulation of the Plasticity Thesis

According to Rawls’s plasticity thesis, the dependency relationship between personal identity criteria and moral theory actually runs from moral theory to the personal identity criteria. Our choice of criteria is a result of our moral theory, not the other way around. As Rawls puts it, the variations between the different criteria of personal identity,

…are accounted for not by this subject but by the distinctive principles and conceptions of the person embodied in the corresponding moral theories. The differences in emphasis arise from the fact that a criterion of identity is tailored to the requirements of a particular moral view. To this extent, the variations among the criteria are not antecedent to moral theory but explained by it.161

This part of the plasticity thesis says our moral theory will determine our choice of the preferred personal identity criterion.

There is also a further part of the plasticity thesis though: Our choice of a social theory and the consequently endorsed criteria of personal identity will also encourage people to nurture to a particular extent what that particular criterion requires. For example, a Kantian view and a Kantian society will “encourage stronger and longer-lasting continuities”162 in persons than a utilitarian society, he writes. The reason for this is that the Kantian ideal is that of “autonomous persons who take responsibility for their fundamental aims over the span of a life.”163 He consequently argues:

There is no degree of connectedness that is natural or fixed; the actual continuities and sense of purpose in people’s lives is relative to the socially achieved moral conception.164

This characterization of the plasticity thesis should be supplemented by a brief look at Rawls’s motivation for this second critical thesis about the relationship between personal identity and ethics. The motivation for it consists in the fact that utilitarians seem to be at an advantage so far. Their view seems to receive support even from theories of personal identity that require very little consciousness. Rawls recognizes this point. As he puts it, utilitarians “can get by with a weaker criterion of personal identity.”165

What he has in mind is the following. Classical utilitarianism uses a conception of the person as a “container-person: persons are thought of as places where intrinsically valuable experiences occur.”\textsuperscript{166} Since utilitarianism does not give intrinsic value to the distribution of goods among these containers, they do not really need to keep track of persons over time. As he puts it a page later, “Assuming that no such experiences have a duration greater than a specified interval of time, one need not ask whether a person having a certain valuable experience in the present interval is the same person who had a certain valuable experience in a previous interval.”\textsuperscript{167}

A Kantian moral theory needs a “stronger” criterion of personal identity, he believes. Kantians use a conception of persons as autonomous pursuer of interests, including long-term interests who accept responsibility over the span of an entire life. For a Kantian moral theory, he argues, “the links of responsibility and contribution have to be traced through time and distribution suitably related to them.”\textsuperscript{168}

In making these remarks, Rawls obviously anticipates the fact that utilitarianism does not necessarily have to rely on a criterion requiring diachronic mental content. He even acknowledges that diachronic psychological relations in persons are often less strong than one might have assumed: “memories fade; hopes are broken; character and will, not to speak of plans and desires, change and sometimes rather suddenly.”\textsuperscript{169} He also seems to acknowledge that the focus on the disintegration within lives in a reductionist theory of personal identity might describe actual conditions better than a “stronger” criterion of personal identity.

\textsuperscript{166} Rawls, “The Independence of Moral Theory,” p. 17.
His way of answering these problems for his view consists in putting forth the plasticity thesis. The actual lack of integration in persons who live under certain conditions tells us nothing about the moral desirability of a weaker criterion of personal identity, he argues. Moreover, he insinuates persons living in a Kantian society will well end up displaying stronger continuities than persons living in a utilitarian society. Thus, he concludes:

…a utilitarian view would be supported by the general possibility of discontinuities only if social theory showed that in the case of other conceptions the requisite connectedness could never be brought about.\textsuperscript{170}

6.2.2 Non-realism about Personal Identity, Realism about Identity Criteria

Before launching into our discussion of the plasticity thesis, it is important to point out some initial common ground between Rawls and Parfit. Rawls and Parfit actually agree in their skepticism about a metaphysically realist understanding of the person or personal identity. They merely disagree about the status of personal identity criteria.

In fact, Parfit has been one of the chief proponents of the idea that the metaphysics of personal identity is not what matters about persons. In his famous branching case he illustrated just that. If two functionally equivalent brain halves of a triplet brother are transplanted into the heads of the two other triplet brothers, and the two brothers end up being psychologically continuous with the first brother, then we are dealing with a loss of the

\begin{footnote}
\end{footnote}
metaphysical one-to-one identity relation, even though psychological continuity will be preserved. Parfit believes what matters has been preserved.

Furthermore, apart from just arguing that the metaphysics of identity is not what matters, he is even skeptical about the existence of an independent metaphysical fact of what we are. If there were a metaphysically robust and independent substance of what we are, then we could technically acquire a new substance on every one of our birthdays without even noticing, which would be rather strange. Also, human beings who lived in entirely different centuries could turn out to have been united to the same substance. This would also be strange, because nobody ever reports things that only a particular person from a different century could have known. Since we subsequently have no evidence for the existence of such independent metaphysical substances, we should remain skeptical about them and form our judgments about persons on the basis of something else.

Rawls would agree with these views. If a person were an independent metaphysical substance, it would be independent of the mechanisms of our societies. The plasticity thesis would have to be false. The conviction that an independent metaphysical understanding of the person is not what our judgments of persons are based on actually constitutes some initial common ground between him and Parfit.

What they seem to disagree about is the fixedness and conceptual independence of personal identity criteria, such as separateness of ends, memory connectedness or character continuity. Rawls endorses a non-realist view of these criteria. For him, they have naturally evolved on the basis of the moral framework that a society endorses. He even believes the societal mechanisms can influence the extent to which people display these criteria. For Parfit, on the other hand, the relevant criteria are independent of societal mechanisms.
6.2.3 Falsifying Plasticity with Conceptual Methods?

Support for the Rawlsian plasticity thesis about criteria can be drawn from considerations against the method of cases in the metaphysics of personal identity. Personal identity theorists tend to probe our intuitions about various criteria of personal identity by constructing hypothetical cases in which an entity persists with some, but not all of the features that we find normal. Usually, this method is thought to tell us what the most central criteria of personal identity are. As Norman Daniels points out, this method might unwittingly assume that there is something fixed about persons that is independent of our normative frameworks.¹⁷¹

If this worry is justified, then a lot of arguments in personal identity theory that come across as entirely metaphysical are actually unwittingly normative. If this is true, it would seem to be a strong argument in favour of Rawls’s plasticity thesis. Even some of the allegedly most metaphysical arguments in favour of particular personal identity criteria would turn out to be triggered by hidden normative assumptions.

Imagine for example a puzzle case that triggers the intuition that we rely more heavily on psychological rather than bodily features when deciding what a person is. There are a number of worries that we may have about the meaning of this result. First of all, the result does not tell us whether the reliance was already a feature of our non-normative descriptive (and synchronic) concept that was merely revealed when we were confronted with the puzzle case. The alternative is upon encountering the puzzle, we just formed an ideal (and diachronic) concept that gives a great (and normative) answer to it.¹⁷² The method of puzzle cases is ambiguous about these two interpretations of their results.

Another problem is that considerations about the normative framework of a society are often specifically excluded in puzzle cases. They are constructed as if we should abstract from those social conditions and give our intuition about the puzzle case completely irrespective of any thoughts relating to our current social framework. For example, imagine we were asked whether a new exotic operation should become a standard practice in health care. The answer to this might well depend on distinctly normative considerations, for example whether the procedure generally allows patients to achieve selfish advantages and avoid unpleasant moral and legal responsibilities. This information might well influence our intuition about the case. Since such social considerations are usually excluded in the puzzle cases, their set-up seems to presuppose that such normative convictions are irrelevant to answering the case.

These are significant worries to the effect that the metaphysical method of puzzle cases is not able to detect the possibly normative character of our intuitions about personal identity criteria and unwittingly suggests that they are not normative. However, we also need to make sure that we do not overstate what these worries show. In fact, they only illustrate that the plasticity thesis could be true, not that it is true. They establish that intuition-probing by appeal to metaphysical puzzle cases is not able to tell us whether the best criterion of personal identity is fixed or a matter of social change. This means that the plasticity thesis could be true, but the opposite is also still open. The best criterion of personal identity might be fixed. The intuitions about puzzle cases might after all reveal what we were committed to all along and might be a genuine reflection of our conceptual convictions rather than the product of social change.

6.2.4 Falsifying Conceptual Analysis with Empirical Methods?
While Rawls insists on the independence of moral theory from metaphysical inquiries, his opponents can insist on the independence of metaphysics and conceptual analysis from
societal mechanisms and moral norms. Rawls seems to assume a form of metaphysical non-realism about the facts that matter about persons. For him, the facts about persons naturally evolve along with the moral framework of their societies. If conceptual realism about those facts were true instead, those concepts would arguably be independent. They would not be subject to historical change and social attitudes.

If conceptual analysis is independent of moral theory, then the improvability problem might run the other way. While conceptual puzzle cases are not able to prove the plasticity thesis right or wrong, empirical methods might not be able to prove a realist conceptual understanding of facts about persons wrong. This raises the question what empirical methods are able to establish about the truth of plasticity or conceptual primacy. If Rawls is right, it should be possible to find some empirical evidence for his plasticity thesis. What kinds of facts would qualify as evidence for it?

Rawls argues that a Kantian moral framework relies on a personal identity criterion according to which persons are “autonomous.”173 They “accept responsibility for their fundamental interests over the span of a life time and who seek to satisfy them in ways that can be mutually acknowledged by others.”174 In A Theory of Justice he also describes them as displaying “separate systems of ends.”175 If it turned out that persons in Kantian societies displayed these features, for example responsibility for their interests, to a stronger extent than the persons in utilitarian societies, we would seem to have significant evidence in favour of plasticity.

175 For example, Rawls, A Theory of Justice, rev. ed., p. 25.
However, there is a problem with this route of inquiry. The choice to observe responsibility behaviour or separateness of ends might itself be a conceptual choice. (As the previous discussions should have revealed, the notion of separateness is also highly ambiguous.) This means it can be argued that any attempt to find empirical evidence for plasticity will itself assume conceptual starting points. One may be skeptical whether empirical investigations with such inbuilt conceptual assumptions can in turn prove that our concept of the person is independent of metaphysical or conceptual assumptions.

Moreover, if the plasticity thesis is true, it has to be true of all personal identity criteria, not just of responsibility behaviour or separateness of ends. In his examples, Parfit often relies on comparatively “basic” criteria. His followers will find it hard to believe that the extent to which people display these basic criteria can change when a society adopts a different moral framework. For example, he often relies on the importance of special self-concern, such as the self-concerned desire to prevent pain for oneself in the future. If the plasticity thesis is true, then the moral framework that a particular society has chosen will influence the extent to which people will want to prevent future pain for themselves. But will people who live in a Kantian society really be less afraid of painful experiences?

Fear of pain is one of the criteria that can serve to distinguish persons from rocks or computer chips. Parfitians believe it is particularly important. The response that fear of pain is not a relevant criterion is not open to Rawls, because the plasticity thesis has been formulated as a general thesis about personal identity criteria. If it should get verified, it cannot apply only to the criteria that Rawlsians prefer. It would have to be true for Parfitian criteria as well.

One way in which Rawls could proceed would be to point out that people are often willing to undergo pain or other unpleasant experiences in order to live up to their moral ideals. If this is the case, then avoidance behaviour in the face of oncoming pain might well
depend on the moral framework one lives in. People in some societies might often be a lot more concerned about fulfilling their moral duties than about avoiding pain. If this is the case, then behaviour or attitudes in favour of avoiding pain would also be subject to plasticity.

However, while this does seem to be a promising route, Parfitians might have a response. As various authors have argued, it is generally impossible to tell for sure whether a particular action was motivated by self-concern or by other-concerned motives. Broad and others argue that this is even true of actions that appear to be particularly selfless, for example in the case of a priest who decides to live among lepers. We will never be able to know for sure whether the priest’s real motivating reason might not have been a self-interested desire for glory and saintliness. 176

If empirical researchers were to question the priest in order to determine whether he was acting out of self-concern, he would likely report to have acted out of altruistic considerations. If he really acted out of a self-concerned desire for glory and saintliness, then reporting that self-concern would defeat the purpose. Even if the priest does not believe he is lying, his reported altruism might have no bearing on the truth. He might simply be unaware of the extent to which he is moved by a desire for glory and saintliness. His unawareness might even be a result of the moral framework that was adopted among the people who were responsible for his social conditioning.

Virtually any action could have both self-concerned and other-regarding motivating reasons. In a society in which self-concern is despised, agents will simply report having been motivated by other-regarding concern. In a society in which self-concern is en vogue, they will report self-concern. This means we would not be able to tell whether the extent to which people display self-concern is fixed or rather subject to change depending on the predominant

moral framework of a society. Hence, the question whether self-concern—with regard to pain or in other respects—is a fixed or a plastic criterion is still unresolved.

6.3 Is the Later Rawls at Odds with the Earlier Rawls?

6.3.1 The Rawlsian Appeal to Separateness Re-visited

To get a full analysis of Rawls’s theses we have to ask whether they might not be at odds with the early Rawls’s claim in *A Theory of Justice*, according to which utilitarianism’s alleged disregard for the separateness of persons will have to lead to its rejection. If there were a significant tension between these two parts of Rawls’s work, then maybe not even all Rawlsians would agree with the theses.

In *A Theory of Justice* Rawls still argues that “utilitarianism does not take seriously the distinction between persons.” It fuses many persons into one and applies the principle of choice for one person to society as a whole.\(^\text{177}\) He also argues, the “reasoning which balances the gains and losses of different persons as if they were one person is excluded.”\(^\text{178}\) In the end he decides:

\[\text{…if we assume that the correct regulative principle for anything depends on the nature of that thing, and that the plurality of distinct persons with separate systems of ends is an essential feature of human societies, we should not expect the principles of social choice to be utilitarian.}\(^\text{179}\)


Is this argument in conflict with the independence and the plasticity theses? The passage can be read as though he thought a particular non-normative property of persons—their separateness—supports the rejection of a particular moral theory. This would be at odds with what the later theses claim.

However, the passage in *A Theory of Justice* can also be interpreted in a way that does not produce a tension with his later theses. On this interpretation, though, the passage is either incomplete to the extent of being misleading, or was not meant to be an independent and further argument for the rejection of utilitarianism at all. Of course the appeal to separateness would also turn out to be an appeal to normative rather than purely metaphysical considerations.

According to the independence thesis, as stated by Rawls, the rejection of a moral theory by appeal to a particular criterion, such as the rejection of utilitarianism, can be denied by endorsing a *competing* criterion of personal identity. Every moral theory, including utilitarianism, can appeal to its own criterion. So if we accept a utilitarian criterion, we will *not* have to reject utilitarianism according to the independence thesis. This does not turn the rejection of utilitarianism in *A Theory of Justice* into a false or invalid argument. It points out that it is a conditional argument though. It is conditional on accepting one possible criterion rather than another.

Furthermore, as the plasticity thesis points out, the separateness of persons might not have been intended as an appeal to a genuinely metaphysical criterion. Instead, it might have been chosen as a *result* of a pre-existing belief or disbelief in a particular moral theory. This means the appeal to separateness might not be an independent additional argument for the rejection of utilitarianism at all. It might merely be a result or a symptom of the normative rejection as such, or maybe simply a different way of stating that we should reject the theory.
6.3.2 Independence and Circularity in a Reflective Equilibrium

Another issue that arises from this is whether the truth of the theses would actually have problematic consequences for the justificatory force of Rawls’s own original position. The veil of ignorance makes certain stipulations about the persons in the original position who are about to choose a normative framework for the society they are going to live in. If the plasticity thesis is true, the moral framework prevalent in our current actual society may have adulterated the stipulations about the persons in the original position. Subsequently, the choices made behind the veil of ignorance might merely reinforce what is actually accepted under current conditions without lending any extra support to a new view.

If the plasticity thesis is true and does not at the same time compromise the justificatory force of the original position, it cannot speak against the conceptual independence of at least some intuitions about personal identity. In other words, the truth of plasticity could not entail that justifications in the other direction—from assumptions about persons to moral theories—are entirely illegitimate. In fact, it would have to entail that justifications in either direction are workable. We could first assume certain central features about persons, then support certain moral theories with those assumptions, and finally, those moral theories would have the power of influencing in turn the central assumptions about persons.

As Daniels argued, viewing Rawls’s original position as a particular feature of a wide reflective equilibrium reveals that this understanding of the plasticity thesis may introduce an independence constraint of a new sort. There are assumptions about persons, procedural justice, social theory and morality as background theories in the way the original position is set up. Then there are the judgments about a moral system that are being made in the original position, which will in turn trigger new views about persons, procedural justice, social theory
and morality. If the original position has justificatory force, the initial assumptions about persons have to be sufficiently independent of the moral judgments they lend support to.¹⁸⁰

If the plasticity thesis is compatible with the justificatory force of the original position, then there may be some common ground between Rawls and Parfit. Both theorists would have to insist that there are at least some assumptions about persons that are independent enough of at least some moral judgments in order to be able to substantiate arguments in which the former are used to lend justificatory force to the latter. The disagreement would merely consist in the initial choice of criteria and in whether some of them might presuppose too much.

This means that Parfitians have two options on how to handle the plasticity thesis. They can disagree with it, but they may also agree with it. First, they can argue that it is unverified or even in principle unverifiable, and that personal identity criteria are really conceptual judgments that are independent of any normative frameworks. Second, they may agree with Rawls that the plasticity thesis is true or at least verifiable, and maintain that the criteria of personal identity that they find important are independent enough of the moral framework that they lend support to. If they chose the latter route, then they will count in their favour the fact that both sides have labeled the utilitarian criteria of personal identity as “weaker” than the ones preferred by Kantians. A stronger criterion is in stronger danger of presupposing aspects of the moral theory it lends support to. This is significant because the comparative strength of the arguments on each side will now depend on who presupposes less.

Conclusion: A Complex Fine-Grained Underdeterminacy Thesis
Broad appeals to the general nature of persons do not seem to be able to establish much for
moral theory. The reason is that there are multiple possible criteria of personal identity all of
which can be given various amounts of weight. Broad appeals to the nature of persons are
simply not specific enough. But even more specific appeals to a particular criterion do not
seem to be able to support very general moral principles such as utilitarianism or justice as
fairness unless they are supplemented by further arguments. Specific appeals to individual
criteria only tend to support particular aspects of such a theory. Such aspects include: the
importance the theory places on subjective welfare; the way in which it concerns itself with
distributive justice; the amount of moral asymmetry it recognizes between the normative
relationship one has to oneself and the normative relationship one has to others; and the
degree to which the theory advocates agent-relativity. This means appeals to the nature of
persons generally underdetermine the choice between very general moral principles. Apart
from this underdeterminacy with regard to global moral principles, the structure between
individual personal identity criteria and particular aspects of moral theories is considerably
more fine-grained than some ambitious endeavours have presumed.

There are various reasons to be skeptical about overly ambitious appeals to personal
identity and separateness. After Parfit’s response to his claims in *A Theory of Justice*, John
Rawls eventually argued that theory selection in ethics is actually independent of one’s view
of persons.¹⁸¹ As the analysis in the last chapter has shown, the best possible reading of the
independence thesis claims there is independence because no moral theory can get ruled out
on the ground that it ignores every criterion of personal identity that has been argued for.
Since no moral theory ignores every criterion that has been argued for, it is also the case that

overly general and unspecified appeals to personal identity can establish close to nothing in moral theory.

Rawls also ends up arguing that moral theories presuppose normative ideals of the person. If personal identity theories are inherently normative, then we are not really dealing with arguments that conclude moral principles from non-normative properties. And if the normative personhood properties from which we conclude moral principles presuppose those moral principles in a significant enough way, then the conclusions are always at risk of being trivial.

There is a similar literature attacking the methods of personal identity theories: This literature says that the kinds of hypothetical thought experiments that are often used to establish the exact nature of persons may also partly presuppose what they seek to establish. As Norman Daniels points out, they might for instance falsely presuppose a paradigm of being immune to any historical and social change. Some of the thought experiments that have been employed to pry different theories apart are indeed so hypothetical that the makers of science fiction thrillers and Hollywood movies have shown more interest in them than any serious psychologists—or ethicists—have shown. Such starting points may be “unwittingly ideological,” as Daniels puts it.

I believe one need not give up on the idea that there are fundamental dependency relationships between at least partly non-normative properties of persons and particular aspects of moral views. The properties in question are non-normative at least in the sense that they do not presuppose the moral principles to which they lend support. Moreover, the point at which appeals to personal identity criteria become normative can be fairly clearly

discerned. It is located at the point at which we assume that the weight we give to the various criteria for the purpose of deciding whether a person has remained identical across time or for the purpose of determining our degree of self-concern should match the weight we give to these criteria in our moral theory. The real question is whether the reasons for our weight distribution in the realm of personal identity and what matters are sufficiently independent from the weight distribution in the moral realm in order for the former to be able to provide independent support for the latter. In this sense, philosophers who presuppose less about their favourite moral theory in their appeals to the nature of persons and what matters will have a more convincing argument in favour of their moral theory.

As argued in the last chapter, it would be very hard to show that we can have absolutely no metaphysical, while at the same time pre-moral, commitment to any particular personal identity criteria at all. Parfit believes that the existence of a metaphysical further fact of personal identity would increase the weight of distributive principles. Its lack, on the other hand, would decrease their weight. Even though it might itself be hard to verify, I know of no argument that has falsified this claim. If he is right about this, then a purely metaphysical, as opposed to distinctly normative fact can have an impact on ethics. This is a point Rawls seems to have overlooked in his argument for independence. Overlooking this aspect has also contributed to the belief that personal identity is an entirely normative matter.

Apart from the role of metaphysical and pre-moral commitments, there are at least three further sources of controversy and confusion in the attempts to understand the relationship between personal identity and ethics. First, there are controversies in personal identity theory that have close to no impact on real life cases. One such example is the bodily transfer debate. The notoriously hypothetical thought experiments of this debate are centred around the case in which a technical device records all mental and bodily states of a particular
person, then initiates the destruction of that person’s body, and orders its recomposition out of
different matter on another planet. These cases managed to pry apart psychological continuity
accounts with an additional requirement of normally-caused physical persistence from those
without such a requirement, but the underlying controversy has little impact on real life cases.

Second, competing theories of personal identity will often give exactly the same
answers when confronted with particular real cases. For example, various criteria can agree
that this author is identical to yesterday’s author. The fact that the competing theories can
sometimes give compatible answers to real life cases has further contributed to the attitude
that we can do ethics without paying attention to metaphysical debates about the nature of its
subjects.

Third, all criteria of personal identity, even ones that are shared by competing theories,
can be given various amounts of relative weight for identity or self concern judgments. How
much weight each feature is given for this purpose can make a real difference for moral
questions though. Most authors seem to have assumed without much ado that a particular
criterion is the most important for our identity or self concern judgments, without addressing
the fact that there are other plausible criteria for this purpose that will support competing
moral views.

Brink, for example, who is no more guilty of this than other authors, assumes that the
kinds of psychological continuity relations we should pay attention to are just the ones that
can apply between separate people—he disregards any others. What he takes to be important
are continuity and appropriate causation of such mental features as beliefs, desires, intentions
and character. Continuity of memory, on the other hand, plays no special role in his account.
No wonder then that psychological continuity does not just seem to support Brink’s self-
extended egoism, but has also been assumed to support the opposite, namely the strictly
individual egoism that Sidgwick had in mind. The latter seems to receive support if we emphasize psychological relations that hold exclusively within individual lives, such as first-person memory. Unwittingly settling on particular ways of fleshing out psychological continuity rather than others will lead to conflicting moral results. In principle one can say that the strangely contradictory moral arguments on the basis of overly general criteria such as “continuity” or “separateness” point at a large-scale equivocation problem that has affected some of the most influential works in modern ethics.

Likewise, it is important to be explicit about the relative weights of different criteria for identity or self-concern judgments. For example, psychological connectedness is a notion that is generally referred to when explaining what psychological continuity actually means. A continuous psychological chain is supposed to be made up of a series of overlapping psychological connections. However, this observation does not tell us yet what the relative importance of the two properties—continuity and connectedness—is supposed to be.

The relative importance of connectedness and continuity for identity or self-concern judgments seems to make a difference for moral purposes. If the exact degree of psychological connectedness is more important than whether or not there is overall psychological continuity, then what matters about persons for ethics will seem less integrated across time. This can lead to different answers when deciding about the relative importance of distributive principles among stages of people’s lives and may also have further moral consequences. If people are seen as naturally disintegrated across time, then applying distributive principles to the different stages of their lives will seem more important. If people are seen as fairly integrated across time, then it will seem more important to apply distributive principles to entire lives. Similarly, if people are naturally disintegrated across time, then
there are limits on how binding a promise will be. If we view persons as rather unified across time, the moral force of a promise will generally last longer.

Other features that can be given different amounts of weight are the particular kinds of connections that matter for identity or self-concern judgments: for example memories, beliefs, desires and intentions. As mentioned above, some of these connections can be mutually caused and shared by separate individuals. For example, we think of beliefs as something that can arguably be caused and shared by other people. Two people can both believe “there will be rain tomorrow” and may have even contributed to each other’s formation of this belief. Other connections can presumably not be shared by separate people, for example introspective awareness of spatial memories. Even the closest friends will not remember the events in the shared past from the same spatial perspective.

Competing theories of personal identity can agree that both memories and beliefs are important in some unspecified way. But the relative amount of weight we give them for identity or self-concern judgments can make a difference for some ethical questions. For example, if shareable beliefs matter the most, then the dichotomy between self and others that underlies some moral theories, such as egoism, will become blurry. Also, moral theories that assign especially large importance to close human relationships will seem to receive additional support. On the other hand, if non-shareable memories matter the most and connections like beliefs matter less, then our views about these moral questions will likely be different.

What we learn from this is that it is important to be very clear about the exact features of personal identity we have in mind and assign value to when launching an argument. The more fine-grained these views, are the clearer will be the relationships between these features and moral views to which they lend support. Only by being rather detailed is it possible to
work out the dependency relationships between personal identity theories and ethics and to
gain further insights into the normative implications of the various possible criteria. To
attempt this, I will now summarize a few of the dependency relationships that have been
established earlier.

*Physical Persistence and Momentary Consciousness*

If our criterion of personal identity merely requires the physical persistence of parts of our
bodies, then the relationship between personal identity theory and ethics might well be one of
rather strong independence. A purely physical persistence criterion can at the most lend
support to a teleological account about the proper functioning of organisms or to an
undifferentiated bias in favour of equality. Such an account can generate some moral
recommendations in the biomedical or environmental realm, but will not have much to say
about a large number of other moral questions that we have.

A physical persistence criterion can get combined with psychological criteria of
various strengths. For example, it can get combined with a requirement for merely minimal
consciousness such as a string of momentary sense impressions. Such a psychological
criterion could not stand on its own, since it does not give us a diachronic account on its own.
Without the physical persistence criterion as a further element, we would not know which
momentary psychological impressions are parts of which “string.” The underlying physical
persistence criterion allows us to give an answer, though.

If we insist on momentary sense impressions as part of our personal identity criterion,
we can generate a moral concern for the momentary subjective welfare of these beings as well
for as an initial bias in favour of equality. This does not amount to full support for any
particular moral theory, even though it seems to be an ingredient for various theories, in
particular for egoism or utilitarianism, but also for patient-centred deontological theories or for some form of egalitarianism.

*Psychological Connectedness*

One kind of diachronic psychological criterion is connectedness. Parfit believes it is one of the aspects that matter for self-concern instead of identity. Even though connectedness is not a transitive notion and can therefore not serve as a criterion of identity on its own, it nonetheless allows us to trace people through time to some extent. In combination with a transitive physical criterion it could serve as a criterion of identity. To be as specific about connectedness as possible one would need to distinguish between connectedness of various possible kinds of mental content. There can be phenomenological experiences and memories, which presumably only hold in intra-personal cases. There can also be connectedness of beliefs, desires, intentions, or character traits, which are presumably shareable by separate persons. The choice between these kinds of connectedness criteria will result in different moral implications. For now I will look at connectedness as a class of criteria.

In general, connectedness seems to give weight to distributive principles, while assigning them to units that are often shorter than entire biological lives. A connectedness unit is that stretch of self-hood time that one is currently still aware of and rather closely related to. This view might therefore support something like egalitarianism of life stages, including intra-personal life stages, and potentially also egalitarianism of strongly connected groups if we believe the relevant kinds of connections can be shared.

A connectedness criterion will always give some special status to those parts of ourselves to which we are strongly psychologically connected. If we believe what matters the most for self-concern is connectedness of intra-personal phenomenological experiences, then
the view can lend support to a form of self-interest with a discount rate that depends on our specific degree of intra-personal connectedness. The condition for this is that we treat connectedness as a matter of degree.

If we believe what matters most is connectedness of beliefs, desires and intentions, and if we also think these mental states can in a significant way be shared and mutually caused by separate persons, then the view will support bias toward closely connected others. It can therefore lead to a high degree of agent-relativity in our moral theory. On this view it would be morally desirable or at least permissible to give priority to those to whom we are closely connected.

_Psychological Continuity_

Psychological continuity is the only relation that seems to support the idea of assigning blame and responsibility even in cases in which a person has irrevocably forgotten the blameworthy action. The reason is that there can nonetheless be a continuous chain of recollections between the time when the blameworthy action was committed and the time when the blame is supposed to get issued.

Giving weight to psychological continuity can also have an agent-neutral impact. It will do so under two conditions. First, it has to be construed as an all-or-nothing relation rather than a matter of degree. If it were a matter of degree, it would support a discount rate instead. Second, the most relevant kinds of mental states that are supposed to remain continuous have to be shareable by separate persons. We cannot give a lot of weight to the presumably distinctly individual mental states such as phenomenological experience memory to reach this result.
First-Person Phenomenological Consciousness

As already mentioned several times, it makes a big difference what kinds of diachronic mental states we are considering when speaking of connectedness and continuity. If first-person phenomenological consciousness and recollections matter the most, and we give a lot of weight to continuity, then we might have some support for a self-interest theory as traditionally construed. It would yield some ground for time-neutral self-bias over the entire course of our lives. If we combine a concern for this kind of mental state with a connectedness criterion, or if we construe continuity as a matter of degree, it will support self-bias with a discount rate instead.

A focus on first-person phenomenological awareness that gives equal importance to all first-person experiencers also seems to support views with strong patient-centred provisions. It can help to ground the duty to aid, and can speak in favour of choosing to rescue the many, even when agent-centred restrictions would prohibit it. It would give us a general reason to improve or protect the experiences of everybody.

Finally, it also goes well with prioritarianism. People’s experiences might be overproportionally worse when they lack a certain good and are already at a fairly low level overall. If we believe this is an intrinsic feature of the good of persons rather than an instrumental consequence of diminishing marginal utilities, then a focus on people’s experience states will give us reason to support prioritarian considerations. These considerations say that it is intrinsically better to help the worst off.

Agency-Related Mental States: Beliefs, Desires and Intentions.

If agency-related mental states such as beliefs, desires and intentions matter the most, and if we believe in addition that these kinds of mental states are not shareable among separate
persons, then we have some reason to support agent-centred side constraints. We do seem to have a particular authority over our own deliberations, decisions and actions that we do not have over the actions of anyone else. This can indicate that we should be more concerned about our own actions rather than anyone else’s.

If we believe these mental states are shareable among separate persons, and we combine them with a connectedness criterion that is a matter of degree, then we will yield some support for an agent-relative moral theory that allows us to give moral preference to significant others. Shared beliefs, desires and intentions would connect us closely not only with other temporal parts of ourselves, but also with significant others.

As the Brink discussion has shown, such a view can also reduce the room for self-other asymmetries in a moral view and will reduce the tension between self-interest and the other-regarding demands of morality. If we give particularly much weight to connectedness of shareable beliefs, desires and intentions, then benefits to significant others will be almost like benefits to ourselves. We would have a non-instrumental self-biased reason to benefit others.

This review is by no means a complete summary of all the points that have been discussed in previous chapters. But it should suffice to support the following more general claims: Some moral views, or components thereof go particularly well with particular criteria of personal identity. At the same time, none of these criteria seem to lend entirely conclusive support to a general moral theory, such as classical utilitarianism or justice as fairness. They only wield such force when combined with further considerations and arguments.

Since the supporting relationships between the various criteria considered on their own do not tend to lend conclusive support any general moral theories, but only to particular
aspects of more general moral theories, I propose calling the relationship between personal identity theory or criteria and moral theory one of underdeterminacy.

Apart from this, the relationship between particular criteria and particular aspects of moral views seems to be more fine-grained than has been presumed by many authors. For example, connectedness of shareable beliefs, desires and intentions as a criterion seems to relate to the degree of agent-relativity that our moral theory should display. Fairly specific criteria will lend some support to individual aspects of a moral theory.

The structure of these relationships is also comparatively complex. For example, a continuity criterion can support long-term responsibility in a Kantian theory, but is also compatible with utilitarianism and may even support utilitarian agent-neutrality. Similarly, as argued in the chapter on consequentialism, a connectedness requirement seems to make distributive principles important. But as a holistic reading of Parfit’s view in Reasons and Persons suggests, it can also be part of a compound criterion that includes continuity and momentary experiences as relevant aspects and that will overall reduce the weight of distributive principles.

Sometimes, the force of a criterion also seems to depend on further conceptual beliefs, for example the force of continuity and connectedness criteria can depend on whether we construe them as relations that allow for degrees or as all-or-nothing relationships. Continuity can support agent-neutrality if construed as an all-or-nothing relation; if it is construed as a matter of degree it can lend support to a degree of agent-relativity. This means we have an interesting array of possible complex relationships between particular personal identity criteria and moral views. There are for example:
cases in which a criterion lends equal support to competing moral views, e.g. a continuity criterion might go equally well with a Kantian theory that wants to assign blame and responsibility across a life time and with agent-neutral utilitarianism

- cases in which the force of a criterion will change depending on whether it is part of a compound or considered on its own, e.g. connectedness of shareable beliefs, desires and intentions will lend support to agent-relativity when considered on its own, but can lend support to agent-neutral utilitarianism when used in a conglomerate with momentary experiences and continuity

- cases in which further beliefs will determine whether a criterion will support moral view x or moral view y: e.g. continuity can either speak in favour or against agent-neutrality depending on whether we take it to be an all-or-nothing relation or a relation that can in principle come in degrees

There might be other interesting complex structures as well. Overall, I believe a complex fine-grained underdeterminacy thesis is an eligible general thesis about the force of appeals to personal identity criteria in moral theory. If this thesis proves to be correct, then we should be looking at the end of the undifferentiated and overly general appeals to the nature of persons that characterized the early days of these disputes.
8 Bibliography


• ------, 1989, "Equal Liberty and Unequal Worth of Liberty, " in Daniels 1089.


Oxford: Oxford University Press.

Oxford: Oxford University Press.

• Scheffler, Samuel, 1979, "Moral Independence and the Original Position,"
  *Philosophical Studies*, 35: 397-403.


• Taurek, John, 1977, “Should the Numbers Count?” *Philosophy and Public Affairs*, 6:
293-316.