I give permission for public access to my thesis and for any copying to be done at the discretion of the archives librarian and/or the College librarian.

_____ _____

Signature Date

Preserving "Simple Suppositions"

A Humean Response to Reductionism about Personal Identity

A Thesis Submitted to the Philosophy Department Faculty of Mount Holyoke College in Partial Fulfillment of the Requirements for The Degree of Bachelors of Arts with Honors.

Summa cum laude.

Lindsay Crawford
Mount Holyoke College
2005

Acknowledgements

I am honored to thank formally my primary thesis advisor, Jay Garfield of Smith College, for his confidence in my work, the breadth of his wisdom and vision, and his untiring commitment to seeing this project reach fruition from its most embryonic stages. I am indebted to Jay for his guidance, patience, and his thoughtful feedback, including both his fairest praise and harshest criticism. Without Jay, I would not have a thesis to defend. Additionally, I thank my second thesis advisor, James Harold, for his dedication and close attention to my project, as well as his helpful comments and advice.

I also thank the University of Colorado-Boulder philosophy faculty and graduate students, who introduced me to Derek Parfit's "Reasons and Persons" and whose help and insight provided a strong foundation for this work.

Additionally, I thank those professors whose influences on my intellectual development have been invaluable: Lee Bowie, Stiv Fleishman, Sam Mitchell, Ann Murphy, and William Quillian.

Finally, I thank my colleagues, Constance Kassor, Kathryn Lindeman, Carolyn O'Mara, and Katia Vavova. These women have set remarkable examples of tireless academic rigor and commitment to philosophy, which have profoundly motivated and shaped the nature of my philosophical interests and pursuits. I am unspeakably fortunate for having found such a strong community of dedicated philosophers.

This work is dedicated to Henry E. Crawford.

Contents

Chapter One: Personal Identity and Reductionism

Thought experiments deriving from recent developments in neuroscience have come to dominate debates about personal identity. One response to these thought experiments, which are often rooted in wildly counterfactual scenarios, is to argue that they are simply so implausible that they are essentially uninformative. Quine writes dismissively, "to seek what is 'logically required' for sameness of persons under unprecedented circumstances is to suggest that words have some logical force beyond what our past needs have invested them with."[1] Here Quine is suggesting that our concept of a person is one that can be explained sufficiently by appealing to our ordinary uses of the concept; to act as if the concept of a person survived in hypothetical cases where we could isolate these ordinary uses would be to misuse the concept entirely and render it unintelligible.

Some, however, argue that these thought experiments are sufficiently plausible such that our intuitive responses to them should be taken seriously. The idea is that by conceptually isolating certain features of our ordinary concept of persons, we can come to understand what the crucial feature is that bases the concept of a person. And deciding what this crucial feature is depends on our intuitive responses to the work these thought experiments do in conceptually isolating certain features.

---

[1] Quine (1972): 490.

Parfit[2] defends this view, and such thought experiments motivate his reductionist analysis of persons. His reductionist analysis has two primary components: a metaphysical view of what matters about persons, and a moral view of what matters about persons. Parfit takes the latter view to be a necessary consequent of the former. So we must note on the outset that there are two distinct claims being made, united by the implicit conditional that if our metaphysical view is correct, it ought to structure our moral view.

In what follows, I will first introduce "My Division," the primary thought experiment that Parfit uses to introduce the problem of personal identity.[3] I will examine in this chapter what Parfit believes to be the problem of personal identity, and how this problem motivates his reductionist claims. I will then explicate Parfit's metaphysical picture, and show the way in which this picture is supposed to bring us to a certain view of persons. Parfit holds that this view of persons structures further claims about agency and morality, which will be unpacked in the coming chapters.

"My Division"

Suppose I have been involved in a near fatal accident, from which I emerge with a badly degenerated body, and the right hemisphere of my brain is rendered irreversibly damaged and utterly useless. Because my body is

[2] Parfit (1984)

degenerating, my heart will eventually stop, and I will eventually lose everything, including my fully functional left hemisphere. The only way I can survive is by transplanting my left hemisphere into the de-brained and receptive body of another person. Let's call the person who receives my fully functional left hemisphere "Lefty." On many accounts of personal identity, Lefty would be my former self. This is because, in the absence of any other options, Lefty can preserve what is believed to be most central about my identity – my brain.

It should first be noted that what gives this thought experiment force is that it has been proven that the two upper hemispheres of the brain can be disconnected, yet still be fully functional. Indeed, there exist cases where people must remove, or are born without, the corpus collosum that connects the left and right hemispheres of the brain.[4]

Parfit dismisses the objection that because we have, as of yet, not been able to transplant one hemisphere of a brain into a de-brained body, that the "My Division" case is impossible and therefore ought not to be entertained. Its impossibility is merely technical. What motivates this thought experiment is the fact that a person's consciousness can be divided into independent streams. This, Parfit says, has been proven by advances in neuroscience.[5] Only if this claim were false would we have a real objection to the use of this

---

[3] Parfit (1984): 255.
[4] Parfit (1984): 254.
[5] Ibid.

thought experiment. If this claim were false, it might be the case that we are, in fact, indivisible substances, like Cartesian egos. And if we were indivisible substances, then of course entertaining this thought experiment would be misleading. But because it has been proven that we can divide our consciousness, the mere technical fact that we have not been able to do exactly what we are describing in this thought experiment is no deep objection.

Now, back to the "My Division" case. If the left hemisphere of my brain lives on in Lefty, while my former body and the rest of my brain degenerate and finally cease to function, it seems that Lefty could appropriately be identified with my former self. Of course, it is no trivial fact that my body and the right hemisphere of my brain do not exist in my continuer, but it seems as though as long as the left hemisphere of my brain continues to function and is able to support human life in a body where there was no brain and, hence, no life at all, Lefty would indeed be my former self transferred into a new de-brained body.

But consider a variant of this case, where both of the hemispheres of my brain remain intact. If we can imagine transplanting my left hemisphere into another body, then we can equally imagine this done twice. Now we have two continuers of my former self – Lefty and Righty.

Parfit argues that there are four possible ways to think of my survival in the outcome of this variant of the "My Division" case: (1) I do not survive;

(2) I survive as Lefty; (3) I survive as Righty; (4) I survive as both. As Parfit argues, none of these conclusions is compelling. Briefly: Conclusion (1) is counterintuitive, because, if we know that our brain survives, even if in a divided form, the survival of the twins with my brain is at least better than death (as Parfit says, "How could a double success be a failure?"[6]); Conclusions (2) and (3) are both arbitrary choices; and Conclusion (4) violates the idea that identity must be a one-one relation.

This case brings two principles of our ordinary conception of personal identity into conflict. These two principles are: (a) No person can be identical to more than one human body; and (b) Whether someone is identical to a future person depends on intrinsic features – it does not matter what is happening elsewhere or what is introduced to the case. Normally these two principles lead to convergent judgments about identity. But because this case involves two people having an equal claim for being my former self, (a) and (b) pull us apart.

What brings these two principles into conflict is that they are joined with the implicit premise in the "My Division" case: (p) Persons lack a unity of consciousness, essential to their survival. Indeed, as Parfit argues, it has been shown that our brains can be split by dividing the corpus collosum, which is precisely what happens for split-brain patients. The separation of the two hemispheres results in a kind of division of labor for each hemisphere.

---

[6] Parfit (1984) : 256.

The significance of this case lies in the fact that we are violating the assumption that persons are unified mental and physical substances – an assumption we take for granted in our ordinary practice of identifying persons.

When cases violate a principle that we normally assume to be essential to persons – in this case, the violated assumption that we are unified mental/physical substances – the question of whether I survive the case becomes indeterminate. That is, we cannot answer the question, "Do I survive fission?" by relying on the assumption that persons survive when they have future continuers identical to them.

Parfit's response to the "My Division" case is, first, it would be impossible to claim that the former person is identical to this person's two continuers, *via* principle (a). Indeed, it would be absurd to claim that both Righty and Lefty are identical to my former self, because they are not identical to each other. Second, Parfit claims that the violation of principle (a) does not mean that there is nothing more to be said about the continuers in this case. Parfit argues that even if we have to abandon (a), we do not have to regard fission as a kind of death. Indeed, all of my former self's intrinsic mental features have been preserved and are supported so as to maintain human life in these two continuers. The fact that these two continuers are not identical to my former self does not change the fact that all of my mental life is being continued and supported in some way.

Parfit here concludes that principle (a) does not matter. Our discourse about personal identity, he maintains, is indeed dependent upon principles (a) and (b). That is, when we talk about identity, we betray a commitment to these principles. But in some cases, these principles are violated, and in these cases the decision to attribute personal identity is indeterminate. In these cases, we ought to abandon the notion of personal identity. This is because we can isolate personal identity from what really matters.

Prior to my division, I cared most about having my mental life preserved. In the division case, my mental life *is* preserved. But in order for my mental life to be preserved, I had to give up the idea of having one future continuer, so as to satisfy (a). Here we see that there is a disjunction between what I care about (my mental life being continued) and what normally supports what I care about (my body as the substratum a future mental continuer). Parfit argues that even if the continuation of my mental life is not supported by exactly one future continuer, my mental life is still nonetheless supported, and this is all that matters. We should not, therefore, regard division as death, even if personal identity seems not to be preserved.

Parfit's conclusion, that personal identity as such does not matter, motivates his reductionism. The account relies on two principal premises. The first premise is (p), which was built into the "My Division" case: The unity of consciousness thought to be essential to the survival of persons does not obtain. This premise joins a corollary of (b): Whatever matters about persons

depends on their intrinsic features. Premise (b) is normally assumed to be secured precisely by the *negation* of premise (p). That is, we assume that the intrinsic features of our mental lives are secured by a unity of consciousness or some kind of essential mental substance. Parfit, however, shows how these seemingly contradictory premises can be dually granted, since premise (b) does not, in fact, depend on the truth of the unity of consciousness claim. Taken together, then, these two premises motivate Parfit's reductionist account.

Parfit's Reductionist Thesis

First, it is important to distinguish the kind of identity about which we are talking when we discuss problems of personal identity. There are two different kinds of identity: numeric and qualitative. When we make the statement, "Ever since his near-death experience, he just hasn't been the same person," we are trading on a distinction between numeric and qualitative identity. We are presupposing that this person is numerically identical over time, but that after his trauma, he is no longer qualitatively identical in some salient way to his former self. We will ask later whether this is coherent, and if so, in what sense.

The philosophical problem of personal identity is posed with respect to the notion of numeric identity. The question, then, "Is X the same person as Y?" is a question about whether "X" and "Y" refer to *one* person. In order to

answer questions about numeric identity, we must have a kind of criterion for personal identity that specifies what a person at one time must have in common with a person at another time in order for these two stages to be stages of the same person.

The idea that there must be a criterion of personal identity presupposes that the question of personal identity is determinate. On this view, when we consider a thought experiment involving some kind of radical change, the question “Am I about to die?” must receive either a “yes” or a “no” answer. Such views are non-reductionist. On these views, there is some identifiable feature of our psychological or physical makeup that constitutes our personal identity. On a Cartesian view, for instance, a person is a mental substance. What matters in personal identity for the Cartesian is the persistence of a non-physical, mental substance that makes each person his own self. We can also imagine a physicalist view according to which there is a further physical fact about us that secures personal identity.

Parfit rejects the non-reductionist view. According to Parfit, there is ample evidence that there is no Cartesian ego (as in cases where the brain can be divided), and absolutely no evidence in favor of this view. Additionally, we have no evidence that there is a certain physical appendage that secures personal identity over time. He argues that if we reject non-reductionism, on the grounds that it is not supported by evidence, we should accept a reductionist view of personal identity.

According to Parfit, a reductionist about personal identity states that, "the fact of a person's identity over time just consists in the holding of 'certain more facts.'"[7] These facts can be described without presupposing the identity or even existence of a person. That is, these facts can be described impersonally. Reductionism by itself does not, however, settle what these facts are, and different reductionists may propose different sets of facts. To deny reductionism, one must argue that while the persistence of a body/brain, and psychological continuity, are both important elements of personal identity, personal identity *consists* in neither of these facts.

A reductionist thesis can take either of two forms: (1) Personal identity consists in physical continuity and (2) Personal identity consists in psychological continuity and connectedness. Parfit argues in favor of a reductionist account of personal identity, according to which personal identity is grounded in psychological continuity and connectedness. Psychological continuity and connectedness constitute what Parfit calls, "Relation R."

Parfit's attention to the psychological constitutive facts of persons does not mean that there are no physical facts about persons. It is just that Relation R is more central to what matters. Persons, on this view, consist in and are nothing over and above certain sequences of mental and physical events.

---

[7] Parfit (1984): 210.

Of course, Parfit insists that his brand of reductionism is not eliminativist. He says that even reductionists cannot deny that persons exist. Reductionists instead ought to distinguish between two kinds of existing. When X exists just because it is constituted by the existence of Y and Z, though X is distinct from Y and Z, X is not a separately existing entity. This is contrasted with the Cartesian claim that a certain mental substance secures personal identity. In the case of Cartesianism, the existence of a Cartesian ego is not only simply constituted by further facts, but it is actually a separately existing entity.

Parfit draws an important analogy between persons and nations, which will be a recurrent analogy throughout this work. Most of us, he claims, are reductionists about nations. We agree that nations exist. We can even distinguish between two kinds of nations: France, a real nation, and Zembla, Nabokov's fictitious nation. But while nations exist, a nation is not a separate entity that exists above the certain citizens, territorial boundaries, laws, etc., that constitute it.[8] So even though we admit that nations exist, it is possible that we could give a complete description of reality by providing an exhaustive account of all of the constitutive facts about nations. Parfit argues that we should regard persons the same way. Though persons exist, we could give a complete description of reality by simply describing all of the

[8] Parfit (1984): 211.

constitutive facts of persons (physical and mental facts) without claiming that persons exist.

Because the reductionist claims to be able to exhaust all of the truths about persons by appealing to the underlying facts to which persons can be reduced, we see now how the claim “personal identity does not matter” is plausible on a reductionist view. The reductionist, then, has no problem dismissing the problem of personal identity in the “My Division” case, because it simply does not matter whether Lefty or Righty can appropriately be identified as being identical to my former self. Instead, all that we ought to care about is the fact that both Lefty and Righty are continuers of my mental life.

Conclusion

In ordinary cases, we can come up with determinate answers to questions about personal identity. Whether someone is or is not the same person s/he was yesterday is a question that can be answered in a number of ways, and reductionism is just another way of answering this question. For the reductionist, personal identity consists in physical and/or psychological continuity. So in most ordinary cases, reductionism and other views regarding personal identity deliver similar verdicts with regard to the identity of individuals.

The reductionist position is clearest when considering problem cases, like "My Division." In problem cases, it is indeterminate whether or not personal identity is preserved, and how it is preserved. The reductionist argues that in problem cases, we can know everything about what happened – that is, we can know every fact about what psychological and/or physical connections continue to or cease to hold, etc. – without answering the question of personal identity.

The reductionist approach to personal identity can be best understood by considering the structure of the supervenience of personal identity on non-personal facts. According to reductionism, phenomena like persons may exist, but they do not exist independently of more basic facts that constitute them.

The reductionist argument has been characterized as the Argument from Below:

> **The Argument from Below:** (1) If reductionism is true, personal identity just consists in certain other facts. (2) If a fact consists in certain others, it is only these other facts that have rational or moral importance. We should not ask whether, in themselves, these other facts mater. (3) Personal identity cannot be rationally or morally important. What matters can only be one or more of the other facts in which personal identity consists.[9]

If we accept the Argument from Below, we ought to direct our attention to the facts constitutive of personal identity rather than the fact of personal identity itself, so as to capture what matters.

---

[9] Parfit (2003): 305.

The important reductionist claim is that persons are deflated constructs. A person is reducible to whatever physical or psychological facts constitute a person. If we were able to know all of these facts, we could completely describe reality without even mentioning the supervening fact of a person. So for the reductionist, "a person, so conceived, is not the kind of entity about which there could be such irreducible truths."[10]

In the following chapters, I will initially grant Parfit's reductionist thesis as a metaphysical claim about persons for the sake of argument. I will challenge Parfit's view that the correct metaphysical view of persons ought to change our views about agency. I will criticize Parfit's view about how our view of agency ought to change once we accept reductionism via a serious consideration of our current moral practices, as they are structured around persons as they are normally construed.

In Chapter Two, I look at what Parfit takes to be the correct view of agency from a reductionist standpoint. To evaluate this in full, I will be explicitly pitting his claims about agency against our commonsense notion of agency. I will then examine Korsgaard's Neo-Kantian response to Parfit that is based upon a rejection of Parfit's picture of agency. We will see, however, that Korsgaard fails to rebut adequately Parfit's reductionist thesis.

In Chapter Three, I will mount a decisive attack on Parfit's reductionist view of persons by comparing his account with that of Hume.

---

[10] Parfit (2003): 296-7.

Some have likened Parfit's reductionist account to Hume's views on persons by calling Parfit a "modern day bundle theorist."[11] I will show, however, that a careful consideration of Hume's view of persons, as it is developed throughout *A Treatise of Human Nature*, will make the crucial differences between their accounts more apparent. I will use my interpretation of Hume's view of persons to bolster a decisive rejection of Parfit's reductionism.

Chapter Four presents a detailed examination of the moral implications of Parfit's reductionist account. I will grant Parfit's claim that a correct metaphysical view of persons ought to structure our moral views, and then assess these consequent moral views. An assessment of these moral views will be structured around the Parfit/Hume debate. What I hope to make clear is that Parfit's moral conclusions are not defensible.

Finally, I will conclude in Chapter Five with a close look at how Hume's position ties in with the modern Minimalist stand toward persons, as it is advanced and defended by Johnston.[12] To this end I will show how Hume's position is reasonable and is echoed in current stances toward the debate of personal identity.

---

[11] Behrendt (2003): 331. The suggestion that Parfit's view of persons is an updated Humean view is one that runs throughout the literature.
[12] Johnston (2003).

Chapter Two: The Reductionist Account of Agency

Parfit argues that accepting a reductionist account of personal identity requires accepting its serious implications for our traditional notions of the self; in particular, for how these notions play out in the areas of moral and prudential reasoning. Reconsider Parfit's analogy between personal identity and nationhood. No one would deny that nations exist, and clearly the concept of a nation is one we use frequently and appropriately. But, Parft argues, a closer examination of what matters when we talk about nations reveals that the facts constituting a nation (i.e., those facts revolving around a nation's citizens and laws) are what matters. Parfit argues that most of us are reductionists about nations. We have no problem maintaining two seemingly paradoxical but compatible ideas here: (1) The existence of a nation just consists in more particular facts, and (2) A nation is an entity distinct from these more particular facts.[13]

Parfit uses the reductionist account of nationhood as an analogy for how we should approach personal identity. Parfit argues that personal identity is to be thought of as a higher-level fact, the supervenience base of which involves more particular facts about psychological and physical connectedness and continuity. Parfit suggests that in virtue of this

[13] Parfit (1984): 211.

supervenience structure, a description of lower-level facts could suffice to explain what matters about persons, since, as personal identity is a supervening concept, personal identity does not add anything important to our description of persons. So, the argument goes, these lower-level facts are what ought to bear on our moral and prudential reasoning, not the concept of personal identity. I will consider whether or not Parfit's nation analogy is apposite to personal identity, and whether it is even correct on its own terms toward the end of this chapter; but for now, let us simply consider how Parfit uses this analogy to argue for what he sees as reductionism's consequences for agency.

The two major consequences of Parfit's reductionist thesis are:

(1) Personal identity over time just consists in more particular facts; and

(2) These facts can be described impersonally, without presupposing personal identity.

Our common sense notions of morality and rationality are structured in terms of personal identity; Parfit hence suggests that revisionary accounts of morality and prudential concern are in order.[14] When we say that personal identity structures these two areas, we mean that we presuppose a unity of consciousness in our moral and prudential reasoning, grounding what I call

[14] It is important to note that Parfit nowhere suggests that, similarly, because the concept of a nation is a separate entity that supervenes on further facts, we ought to revise our discourse on nations. We will return to this issue later.

"common sense egoism." Common sense egoism is a theory about rationality and prudential reasoning. The common sense egoist believes that it is rational for an agent to do things in his own interests, and that it is rational for an agent to be specially concerned about his own future in virtue of its being his. One theory that buttresses these claims is the Kantian notion of a unified consciousness that allows us to unite different experiences in virtue of the fact that we are unified agents over time. In this chapter, I will focus on how the Kantian notion of a unified consciousness supports the claims of the common sense egoist, in order to bring to the forefront a more defensible response to Parfit's reductionist account, as well as to see whether the Kantian defense is a strong one.

Reductionism undermines the common sense egoist's position as well as corresponding views in the domain of morality. In its emphasis on the importance of the psychological relations that obtain between different stages of a person's life over the unity of a whole life, and the ways in which these relations hold to different degrees, reductionism effectively rejects the common sense egoist's view that it is rational to structure our prudential concerns around temporally extended persons. Once we shift our concerns from persons to degrees of psychological connectedness, it is suddenly incumbent upon us, according to Parfit, to construct and incorporate new prudential/moral units into our reasoning to reflect this shift in concerns. So, it

is clear that once the reductionist has undermined the common sense egoist's position, as it specifies what concerns are rational to have about ourselves, this will have important consequences for our moral reasoning.

In what follows, I will examine the theoretical merits of Parfit's reductionist account; I am bracketing a review of the practical consequences of applying reductionist principles until Chapter Three. Here, we will consider Parfit's account of how reductionism can change our notions of prudential and moral concern. After considering what reductionism entails about agency in general, I will turn to the work of Kant and Neo-Kantians who defend the view that the unity of consciousness is central to personhood and to our discourse about agency. I will explicate Korsgaard's argument that Parfit has failed to take into account a particular feature of agency, which Korsgaard argues is our ability to claim authorship over our actions and our futures. This account, Korsgaard argues, is unavailable to Parfit, because Parfit views persons from an essentially theoretical perspective. Parfit's perspective is theoretical because it emphasizes the extent to which the lower-level facts that constitute personal identity capture the truth about our notions of agency. Ultimately, I will argue that Korsgaard's account of personal identity amounts only to a counterargument, which does not properly rebut Parfit's reductionist claims. As we will see, Korsgaard's emphasis on the practical perspective, which she claims we *must* take toward our actions, can be easily undermined

by the reductionist. The reductionist needs only to reject Korsgaard's claim by appropriately arguing that the supposed *necessity* of this stance does not entail its reasonableness or normative force.

Common Sense Egoism

Practical reasoning involves evaluating a number of potential actions, and choosing to take one. Our decision to choose one action over another reflects the action's expected utility. To make a rational choice means to choose the action with the highest expected utility.[15] Theories of rationality differ in two ways central to our present discussion. These theories differ in terms of how benefits ought to be allocated across persons and across times. The theory of rationality with which we are primarily concerned here is common sense egoism. Common sense egoism represents our basic beliefs about ourselves and our futures.

Underlying common sense egoism is the claim that these beliefs structured around personal identity have *non-derivative value*. In other words, the common sense egoist claims that our concerns about ourselves and about our futures are structured in terms of personal identity, and that their *prima facie* reasonableness is their defense. For example, the common sense egoist would argue that one does not have to cite reasons to justify self-referential

[15] For my purposes, I am construing "utility" in a broad sense. "Highest expected utility" should be understood as the best outcome of an action, either for the agent, for others, etc.

concern. We hold basic beliefs that presuppose persons as natural entities of concerns, and these concerns do not have to be justified further. As Johnston, who appeals to the non-derivative warrant of self-referential concern, argues, a defense of these basic beliefs would be based on the fact that we find them natural, and that, so far, critical reflection has not shown these beliefs unreasonable.[16]

The claim that certain self-referential beliefs have non-derivative warrant can be understood as follows. Certain basic concerns constitute a complex pattern of other concerns, and these concerns are justified to the extent that they accord with other concerns and that they stand the test of informed criticism. For example, consider the general belief that many of us have that the world ought to be a better place. How is one supposed to justify the basic reasonableness of this claim? It seems that when we try to justify certain claims, our justification is based on an appeal to simply another claim. This pattern of justification must finally end based upon a claim that we presume to be natural and reasonable. And to discard these basic beliefs all at once would be to undermine an entire pattern of beliefs and concerns of which these basic beliefs are a central part. So, based upon the notion that certain concerns and beliefs, like love and self-referential concern, can have non-

[16] See Johnston (2003): 268-9, for a more detailed account of non-derivative self-referential concerns.

derivative warrant, common sense egoism's substantive claim is that we are each distinct persons, and that the boundaries between persons matter.

Common sense egoism echoes some of the core claims of what Parfit calls the Self-Interest Theory (S). S states: "For each person, there is one supremely rational aim: that his life go, for him, as well as possible."[17] S differs from common sense egoism only in that it does not explicitly appeal to the concept of non-derivative concern explained above, and is a broader base upon which to rest more specific claims that reflect the idea that we are separately existing persons, and that this separateness matters. It is important to note the parallel between the two in order to understand Parfit's theoretical account of reductionism, which comes as a direct response to S. Common sense egoism helps us in a way that S does not, in that it provides a basis for considering the Kantian claims we will explore in this chapter.

S-theorists and the common sense egoist defend two claims that Parfit uses to structure his discussion of reductionism. The first claim is:

(1) A person has reason to make sure that the actions he makes benefit him, in virtue of the fact that these actions are his.

This is an *agent-biased* conception of rationality, for it is implicit that it matters to whom the benefits of an action are accorded. S, in this fashion, assigns value to the actions that most benefit the person performing an action.

[17] Parfit (1984): 4.

Opposing theories hold that it is rational to take actions that distribute benefits across persons. Such theories are altruistic, for according to these theories, it is of no moral significance to whom the benefits of actions are made. Altruistic theories are united with theories like S, however, in that, according to both, it is morally important that the way in which the harms or benefits of an action are distributed across persons matters, and ought to factor into how we act.

Altruistic theories do not claim that we can attach equal weight to everyone's interests. Indeed, altruistic theories do not assert that there are no significant differences between persons. A theory that truly disregards the separateness of persons would be *agent-neutral*, whereby our actions should not concern anyone's welfare *per se*, but rather should be evaluated on the basis of the type and magnitude these benefits yield.

The second claim S makes is:

(2) A rational agent has reason to be concerned about his future selves, in virtue of the fact that his future selves are temporal stages of one and the same person.

This does not mean, however, that a rational agent cannot apply a kind of discount rate with respect to prudential concern to his future selves.

One of the ways Parfit rejects S is by rebutting what Parfit calls S's temporal-neutrality claim (which states that a rational agent is equally

concerned about all of the temporal parts of her life). Parfit says that for the S theorist, "the force of any reason extends over time. You will have reasons later to try to fulfill your future desires. Since you will have these reasons, you have these reasons now . . . What you have most reason to do is what will best fulfill, or enable you to fulfill, all of your desires throughout your life."[18] Parfit argues that this claim is implied by the S theorist's explicit claim that we are unified persons over time, and, as such, any future temporal stage belongs to us in an important way.

Parfit presupposes that the only way we could defensibly care less, or in a different way, about our futures than we do about ourselves at the present stage is if the temporal stages of our lives were somehow demarcated so as to reflect justifiably a change of attitude toward them. Parfit takes it that as long as the S theorist continues to hold that all future stages of himself belong to him (i.e., an agent is a single temporally extended entity), the S theorist must admit that there is no way of properly demarcating one temporal stage from another. Thus, the S theorist must make the counterintuitive claim that our concern for our futures must be equal in force to our concern about ourselves presently.

Parfit also notes that accepting temporal neutrality and agent relativity concurrently is inconsistent. "S allows the agent to single out himself, but

[18] Parfit (1984): 137.

insists that he may not single out the time of acting. He must not give special weight to what he *now* wants or values. He must give equal weight to all the parts of his life, or to what he wants or values at all times."[19] Parfit challenges what he sees as an indefensible asymmetry in the S-theorist account of persons: While the S theorist holds that we have reasons to be agent-relative, the S theorist concurrently holds that we must be temporally neutral. For Parfit, the asymmetry at work here is that the S theorist accords a special status to the agent, but denies this status to the time of acting. Denying a special status to time, while at the same time granting a special status to agency, is an inconsistency in S and is indefensible. Indeed, Parfit points out that being fully temporally neutral would require the S theorist to care equally about all *past* events as he does present and future events. This, Parfit claims, is clearly untenable. Therefore we ought to reject S on the grounds that it is a hybrid theory and accords importance to agents and time arbitrarily. [20]

The point in bringing common sense egoism into the discussion is not to avoid Parfit's pressing criticisms against S. Instead, common sense egoism has been proffered as a view that retains the vital claims that motivate S, but which is not committed to Parfit's dubious inferences. It should be noted that Parfit's S seems to be a weak straw man whose eventual rejection Parfit paints as a point in favor of the acceptability of reductionism. At this point I will

[19] Parfit (1984): 140.
[20] Parfit (1984): 193.

sketch common sense egoism as it diverges from S in its explicit rejection of temporal neutrality.

There are two immediate red flags that we should note when Parfit sketches S's temporal neutrality claim. First, it is clearly false to characterize any view under the "common sense" heading when it is dubious whether even one person is wholly neutral toward all the parts of his future. Oddly enough, however, Parfit argues that S is a common sense view – in fact, one held for millennia.[21] In fact, most people treat their future selves as distant enough to warrant less concern about them than their present selves.

Second, Parfit presupposes (falsely) that to have agent-biased self-concern, one *must accept* temporal neutrality. This is implicit in Parfit's sketch of S. On the common sense view, however, an agent can be specially concerned about his own life in virtue of its being his, but also be justified in having more concern for his present self than his future selves. Of course, an agent can feel quite distanced from a future self and have this distance reflected in his concerns, but still regard these futures selves as *his*. It is not clear why the S theorist should be committed to such a strong claim simply on the basis of the premise that an agent's rational aim is for his life to go as well as possible.

---

[21] Parfit (1984): 130.

It is important to appreciate the connection between the common sense egoist's two commitments. A crucial reason for believing that personal identity is especially important is that we believe, in some sense, that our futures belong to us in a way that they cannot belong to anyone else (even though, contrary to what Parfit claims, we are not *consequently neutral* with respect to every stage of our future selves). For instance, I can be extremely fearful of and even make substantial personal sacrifices to prevent the future pain of a loved one, but this anticipation does not have the same fearful quality that the personal anticipation of my own future pain has. Our belief that we-now are identical to ourselves at future temporal stages goes hand in hand with the egoist's claim that it is rational to have our actions benefit ourselves. It is because a person at one stage of one's life is the same person at a future stage that it is rational for this person to do actions that might have future benefits, because the recipient of these benefits is, quite simply, the same person who performed these actions.

Even jointly, however, these commitments do not entail that we are *equally concerned* about the temporal parts of our lives simply because personal identity matters a great deal. Indeed, we do have reasons to be concerned about our future selves, but this does not necessarily lead to the claim that we have to be equally concerned about every stage of these selves.

Reductionism and the Unity of Consciousness

Parfit defends reductionism firstly by undermining the idea that we are unified conscious subjects over time. Reductionism offers two ways to understand the concept of the unity of consciousness: First, the unity of consciousness is not a necessary metaphysical condition for agency. Second, the unity of consciousness, as it fails to be metaphysically necessary, consequently fails to capture what matters in moral and prudential reasoning.

The claim that persons possess a unity of consciousness over time is one that has non-derivative warrant for the common sense egoist. The common sense view is that an agent is a unified "subject of experience," and these experiences are united in virtue of being had by one person. Parfit argues that appealing to ownership to explain the unity of consciousness is insufficient. Indeed, he argues that it is possible for there to be more than one "subject of consciousness" within a single person. Appealing to the fact that these unities are possessed by the same person does not explain the phenomenon of unified consciousness. Parfit defends this position on the basis of developments in medical science that have proven that the two upper hemispheres of our brain can be disconnected, causing two separate spheres of consciousness. As we saw in the previous chapter, this evidence motivated the "My Division" case.

In actual cases, people whose hemispheres have been divided can have different experiences in each hemisphere, without the one hemisphere knowing what the other hemisphere is experiencing. Our left hemispheres control our right arms, and vice versa. So, it has been shown that something presented in the right visual field (which corresponds to the left hemisphere) can be recorded when the person is asked to write with his right hand what he has experienced. For example, a person with a divided mind is shown the colors red and blue, each in one visual field. When asked to record what color he has seen, by writing an answer with each of his hands, one hand will write "Blue" and the other, "Red."[22] Parfit takes it that cases like these support the idea that appealing to ownership to explain the unity of consciousness is incorrect in light of the possibility that a person's two hemispheres can entertain two different modes of consciousness.

In accord with these findings, Parfit has us imagine a counterfactual scenario, "My Physics Exam."[23] This case is motivated by the conclusions of the "My Division" case. Suppose there is a woman whose two brain hemispheres are exactly alike in ability, and that she has the ability to divide her brain at will with the raise of an eyebrow. She is taking a physics exam and must finish the last question on the exam in the remaining time. She foresees two possible ways of tackling the problem. In order to see which way

[22] For further discussion of split-brain patients, see Parfit (1984): 245-6. Also, see Nagel (1979).

is better, she divides her mind for the 10 minutes and each hemisphere goes about solving the problem. How should we regard this case? Once she has reunited her mind, she seems to remember having done both methods for the problem, but when she was doing each of them, she was not aware of the two methods' concurrence. So in the Physics Exam case, we cannot unite the experiences being had in each hemisphere by appealing to the fact that there is one person who is the subject of both experiences. We cannot explain the unities in each hemisphere by claiming that I am experiencing all of these experiences, because it is possible for me to be unaware, in one hemisphere, of what is going on in the other. To argue that these experiences are all being had by me suggests that the two hemispheres are identical, which is clearly incorrect.

Still, Parfit suggests, one might just take this case to show that instead of there being a single subject of experiences, there are two. On this view, the unity of consciousness in each hemisphere is explained by ascribing to each hemisphere a "subject of experience." But when one begins to talk about multiple subjects of experience, one can no longer claim that subjects of experience are *persons*. Indeed, the physics exam case shows that there can be supposedly two "subjects of experiences" that are not identical to each other, but which exist in a single person. So this view is committed to the idea that

---

[23] Parfit (1984): 247.

the life of a person can involve different “subjects of experiences,” and that these subjects are not persons. But, Parfit argues, talk about “subjects of experiences” wholly different from persons ought to raise skeptical hairs. Indeed, the idea that a person is what unites several experiences at a single time motivated the appeal to “subjects of experience,” but now it appears that in order to maintain the concept of a “subject of experience” in light of the Physics Exam case, we must to *abandon* the idea that subjects of experiences are persons.

The reductionist denies that we have to call each hemisphere a different “subject of experience.” Instead, we have only the metaphysical facts of the matter. In this case, there are only different states of awareness at a time, each consisting of different experiences. My mind is divided, then, only because there is not a single state of awareness. It is important to note in the reductionist response that this is not a redescription of the problem we had earlier. To claim that each hemisphere is a “subject of experience,” and to claim, instead, that each hemisphere is merely a state of awareness, are two very different claims. The first claim presupposes that it is a deep fact that the experiences in one hemisphere are united. In effect, the first claim leads us into a discussion about agency, since those who hold this view claim that to explain why experiences are united, one must appeal to personhood. The second claim, on the other hand, requires the denial that there is anything deep

involved in the unity of consciousness in each hemisphere. The facts still stand: the woman in the physics exam case is computing a problem in two different ways in each hemisphere simultaneously. The reductionist just says that once we have all of the facts regarding what is happening in each hemisphere, we have exhausted the issue.

So, Parfit argues that the unity of the experiences in each stream cannot be explained by an appeal to persons. There are two alternatives. We can either call each hemisphere its own subject, or we can take the reductionist route and deny that ascribing experiences to a subject matters *over and above* the psychological facts of the matter.

Implications of the Reductionist View

For the reductionist, all that matters are the psychological facts about a person. It is not important to describe these facts as properties of a subject. In some cases, like the physics exam case, it is simply false to say that there is a single subject of experience, and that this subject is a person. And while we can press on to allow for multiple non-person subjects, we have to accept that subjects of experience are not what we originally had in mind (namely, persons) when we thought this appeal was important. If we accept this, we then see that “subjects of experience” is just a label for the facts of the matter. And this is just another way of describing reality, and this description does not

in itself matter. Thus, Parfit says, we can discard the label "subject of experience" and speak directly about the metaphysical facts. And this can be done impersonally, as the reductionist does in response to "My Physics Exam."

As a matter of convention, we ascribe thoughts to thinkers, actions to agents, etc. In this sense, the reductionist argues, it is true that thinkers and agents exist. But agents and thinkers exist only in virtue of *the way we talk*. If there were a metaphysical entity that was a necessary and sufficient condition for personhood, like a Cartesian ego, then the concept of personhood might have real import. But there is no Cartesian entity, and instead, all we have are the psychological facts that constitute a person. Here we should note that Parfit assumes that if a fact consists solely in certain others, it is only these other facts that have rational or moral importance. Psychological facts ground our concept of persons. This is the basis of Parfit's supervenience claim about personal identity. We will call this claim into question in Chapter Three.

In some cases, like the physics exam case, it is an indeterminate matter whether or not one should properly say that there are two subjects of experience or one. These indeterminate cases show us that even if there are normal cases where our concept of personhood remains intact, the mere fact that we have been able to preserve our concept is trivial. As such, Parfit suggests that we can "redescribe any person's life in impersonal terms. We

could describe what, at different times, was thought and felt and observed and done, and how these various events were interrelated. Persons would be mentioned here only in the descriptions of the content of many thoughts, desires, memories, and so on."[24] For Parfit, the concept of personhood is constituted wholly by more particular psychological characteristics. And in describing all of the psychological facts of the matter, we do not need to reference persons, because the concept of personhood is merely a label.

The reductionist move to deny that persons are anything deeper than their constitutive facts has important theoretical consequences for prudential concern, as prudential concern is structured around the notion of personal identity. We can now see how the reductionist and the common sense egoist come into conflict. The common sense egoist holds that we have reason to be concerned about ourselves and about our futures. Our concerns about ourselves and our futures, as they are structured around personal identity, are utterly natural ones. Parfit argues to the contrary. If it is possible to redescribe a person's mental states without ascribing these psychological facts to a subject, Parfit argues, it is less plausible that personal identity ought to be the principal guiding tool for prudential reasoning.

[24] Parfit (1984): 251.

How Reductionism Bears on Prudential Concern

Prudential concern is structured in terms of personal identity. When I consider the future, it is my future about which I am concerned in a special way that I am not concerned about the futures of others. This is because there will presumably be a future person identical to my present self, whose welfare matters simply in virtue of this identity relation.

There are two options open to the reductionist with respect to what should happen to our conventional notions of self-referential concern. First, one could claim that since reductionism is true, we have no reason to be concerned about our own futures. This is because the only thing that justified being especially concerned about our futures was personal identity (or, more explicitly, the fact that I-now am identical to all of my future selves). This is what Parfit calls the Extreme Claim.[25] The common sense egoist might claim that this is the necessary consequence of accepting reductionism, because personal identity, on the common sense egoist view, is all that matters as the basis of prudential concern. So when we abandon the notion of personal identity and resort to talk only about Relation R, we cease to have any reason to be concerned about our own future.

[25] Parfit (1984): 307.

Parfit notes, however, that the Extreme Claim is not the only possibility open to the reductionist. Parfit suggests another possibility – one which allows us to have future-directed concern, but concern not structured around personal identity. This is the Moderate Claim: *Relation R* gives us reason to have special concern about the future.[26] The Moderate Claim gives us reason to be specially concerned about our future, but at a price: because psychological connectedness between two temporal stages of a person can hold to different degrees, one cannot defensibly claim that one has reason to be equally concerned about all of the parts of one's future. Rather, one has reason only to be concerned about oneself at different parts inasmuch as these are strongly psychologically connected and/or continuous. For example, consider yourself now and yourself 50 years hence. While you-now and you-50-years-future will presumably be continuous with each other, there will be little to no psychological connectedness. And "since connectedness is one of my two reasons for caring about my future, it cannot be irrational for me to care less when there will be much less connectedness."[27]

Parfit urges us to consider parts of our future as being ours at a "discount rate."[28] This discount rate concerns the weakening of connectedness, and this weakening gradually deepens in degree over time. So

[26] Parfit (1984): 313.
[27] Parfit (1984): 315.
[28] Parfit (1984): 314.

it is rational to care more about our nearer future selves than our distantly related future selves.

It should be clear that the reductionist's Moderate Claim is a rejection of the common sense egoist's appeal to persons as they are normally construed as being the appropriate units of prudential reasoning. The Moderate Claim proposes that we shift our concern so as to focus on the psychological facts of the matter, i.e., the degree to which one temporal stage of a person is connected to another. The conflict stems from the fact that the common sense egoist holds that our person-structured concerns have non derivative warrant, and so, the concept of persons, even if constituted by further psychological facts, has its own import. The reductionist denies this by appealing to the supervenience claim that if a fact just consists in more basic facts, it is only these more basic facts that matter.

The structure of prudential concern has crucial implications for the structure of morality. So, if we accept either the Extreme Claim or the Moderate Claim about prudential concern, we will also have to accept similar claims about how we should restructure moral reasoning. On the common sense egoist picture, moral reasoning ought to be principally structured around persons. That is, the more specific views we have in different realms of moral reasoning (such as, for example, distributive justice and punishment) ought to reflect the general notion that persons are unified, temporally extended beings

and that the separateness of persons matters. I will return to these topics in morality specifically in Chapter Four. In what follows, I will show what reductionism implies about agency in general, which will serve as a foundation for some of the later discussions in Chapter Four.

Beyond Prudential Concern

On the reductionist view, once we have undermined the claim that persons are unified subjects over time, as well as the claim that personal identity matters, the distinctions between prudential concern and moral concern are not quite as definite. Indeed, the importance of the separateness of selves (i.e., that it was a non-trivial fact that one person was a single agent different from other agents) helped solidify this distinction, since prudential concern is defined by my beliefs about my own life, and morality, on the other hand, is defined by my beliefs about others. But, as Parfit showed in the Physics Exam case, the label of personhood is not a deep fact. Instead, all that really matters are degrees of psychological connectedness between temporal stages of a person. So, in the same way that the reductionist can either reject the claim that we should have any special concern about our own futures (the Extreme Claim) or adopt R-relatedness concern (the Moderate Claim), the reductionist can apply these two claims to morality.

The Extreme Claim about morality in general is, analogously, that we have no special reason to be concerned about the entities (i.e., persons) that we were formerly concerned about when we operated under the common sense egoist view. Alternatively, we could adopt the Moderate Claim. The Moderate Claim about morality just states that we can shift our concern from persons to R-Relatedness.

The reductionist view, Parfit claims, undermines the boundaries between lives and the unity of agency over time to the extent that prudential reasoning can become, in fact, moral reasoning. It may help if we examine a case that Parfit uses to flesh out how following the reductionist line blurs the distinction between moral concerns and prudential concerns. Consider the following example. It is a common belief that it is imprudent for an agent to act in a way that will negatively affect her later in life. Consider smoking. Suppose the agent is a smoker at age 20. She knows full well but cares little about the health problems that smoking may cause her later in life. One reason to justify this disinterest in the consequences of smoking is that she does not identify with this future self who will pay the price. For the reductionist, it may not be irrational to smoke now, simply because the future self who may suffer the consequences of her smoking now is so weakly psychologically connected to her-now that she is justified in not caring about this future self's welfare. This is the Extreme Claim.

Parfit argues this is a possibility. But accepting the Moderate Claim is just as much an option open to the reductionist. According to the Moderate Claim, we ought to shift our concern to reflect R-relatedness. Once we accept that R-relatedness is all that matters, we must accept the idea that one can have future selves very weakly connected to oneself now. If this connection is sufficiently weak, person-now can be justified in not identifying with this later person, and effectively think of this later person as someone else.

Once we have accepted the idea that future selves can be so weakly connected to a present self that deeming this future self a different person is reasonable, it seems plausible that our attitudes toward this future self ought to be moral considerations rather than prudential concerns. If we accept reductionism in the former example, we could call the smoker's imprudence *immoral.* The 20-year-old smoker can be thought of as carelessly hurting someone to whom health concerns resulting from a smoking habit will directly affect, and so the reasons for her not to smoke can be expressed in moral terms. On the reductionist view, then, the smoker can no longer excuse her smoking habit by arguing, "I'm only hurting myself, so you cannot intervene," because she cannot appeal to the unity of her temporally extended life to justify her actions to herself.

Agency from the Reductionist Standpoint

As we have seen, Parfit maintains that all that matters are the metaphysical facts about persons, not persons themselves. In the Parfitian sense, ascribing experiences or actions to an agent *qua* person is just a linguistic exercise. The reductionist admits that, in some sense, persons exist, but that persons so-called are just convenient constructs constituted by more particular facts. For the reductionist, it is still true that persons perform actions and are the recipients of others' actions. But what matters in moral and prudential reasoning is the quality of these actions and not necessarily to whom or when they apply.

Return to what we identified in the onset as the two major consequences of the reductionist thesis. These are:

(1) Personal identity over time just consists in more particular facts; and

(2) These facts can be described impersonally, without presupposing personal identity.

Parfit argues that once we accept (1) and (2), we have exhausted the concept of persons. Indeed, because personal identity is an upper-level supervening fact, it is possible to describe sufficiently all of the lower-level facts of the matter and describe these in such a way that we do not need to reference personal identity at all. The attribution of the concept of a person can be made after we have described all of the lower-level facts, but Parfit

argues that we need not do this additional work to get any closer to understanding the fact of the matter.

Implicit in Parfit's arguments is the view that persons are entities to which we can ascribe a number of different psychological and physical characteristics. On this view, it may seem plausible that our concept of a person could be fully satisfied if we were to have a fully exhaustive list of these empirical attributes. However, it is not obvious that this is the only view by which we can understand persons.

Kant famously argued that we can consider persons from more than one perspective. A discussion of Kant's distinction between the phenomenal and noumenal views, in regards free will, will serve as a starting point and a helpful analogy that will guide Korsgaard's discussion, as Korsgaard's position depends upon this Kantian insight.

Kant's Two Perspectives View

Kant argues that we view ourselves from a theoretical standpoint, where we become natural entities whose actions and futures can be predicted and causally explained. On the other hand, we can see ourselves as agents, for whom choices are real and our futures dependent upon how we shape them. These views, when taken together, may appear inconsistent, but this is not an argument against seeing persons from these two views. There is nothing

contradictory about allowing our views of persons to be radically different from one another, precisely because the two perspectives that motivate different views of persons reflect different relations persons have to their actions.

From the theoretical standpoint, we understand our actions to be predictable and largely out of our direct control. This is because we can see the entirety of our lives as composed of various actions causally related to later actions, and from this standpoint, persons appear as entities whose activities could be wholly explained by a sufficient account of our mental lives and the history of our actions. From the practical standpoint, however, we see ourselves as being confronted by pressing choices that we must take on. We have no alternative but to act in the face of choices, and so our relation to our actions, from this perspective, appears to be one that is distinguished by a sense of authorship. The need to take action in the face of choice makes it necessary that "every rational subject [be] a law-making member in the kingdom of ends; for otherwise he would have to be regarded as subject only to the law of nature – the law of his own needs."[29]

Now this is not to say that, from the practical standpoint, our sense of authorship and agency are mere illusions. This would be to assume that the theoretical standpoint in some way undermines the practical standpoint,

[29] Kant, (G 439:85)

perhaps because it offers us an objective, empirical explanation of our lives and our behavior in a way that the practical standpoint cannot. But this would not respect the fundamental independence of each perspective, and the differences between them in regards to their utility. Rather, it is because we must make choices that our attitude of authorship toward our actions is a necessary attitude to take, regardless of the metaphysical facts.

This claim about the different perspectives we can take toward persons is perhaps clearest in Kant's discussion of free will. From a theoretical standpoint, it may appear that we are all determined beings working in accordance with laws beyond our control. Indeed, from the theoretical standpoint, we exist in such a way that our interactions can be wholly explained and determined. Upon theoretical inspection, then, it appears that no there is no free will. But, Kant argues, this does not mean that there cannot be another sense of free will that survives theoretical inspection. In order to retain some concept of free will, all we need to do is turn our attention toward the way in which we feel ourselves to be agents, acting in our own interests and as the authors of our intentions. This is the practical perspective. From the practical perspective, we see that there are options open to us, and we must necessarily choose one option among others. From this very act of choosing we may consider ourselves as having a free will.

The disjunction Kant points to between theoretical and practical reason shows that we can assume more than one attitude toward ourselves as actors. We will now examine how this claim motivates Korsgaard's rejection of Parfit's reductionist claims. As we might anticipate, Korsgaard will use this approach to argue in favor of an account of persons that survives Parfit's reductionist claims. So while it may be true that from the theoretical perspective, we are simply constituted by lower-level metaphysical facts, there is the practical perspective from which we *must* regard ourselves as agents in order to make decisions about what we ought to do and how to act.

Korsgaard on the Unity of Agency

Korsgaard works from the Kantian picture of the unity of consciousness and is motivated by his idea that empirical facts do not necessarily bear on the reasonableness of practical reasoning. Korsgaard argues in two ways. First, she insists upon the necessity of the unity of agency. The unity of agency can be understood as the authorial relationship we have to ourselves and our future selves. Korsgaard argues that the unity of agency underlies the unity of consciousness. Second, she argues that unity of agency is justified in so far as it is a practical necessity. So for Korsgaard, the

unity of agency explains how our actions and choices are different from mere behaviors, and why this has normative significance.[30]

Korsgaard argues that Parfit's mistake is precisely that he focuses on the theoretical perspective and ignores the possibility of the practical perspective we can take to persons. The practical perspective includes within its scope an account of the relationship one has with one's future selves and actions. This relationship is left out in the reductionist approach, since all that matters to the reductionist are the facts of experience and actions, not how we ascribe these to agents.

Korsgaard first asks the reductionist what reasons he has to think that he is an agent right now, disregarding the Parfit's problem about identity over time. What makes the I-now identical to I-now? Korsgaard argues that our agency is grounded by a kind of unity with which we approach our concurrent desires and interests. The fact that I am a single person right now allows me to explain how it is that I can resolve various conflicting feelings in one place, such as hunger, my visual experience at the moment, the pressure on my left foot, etc.

When we consider what makes us the same person over time, she argues, we see that our desires and projects develop over time and interrelate, and involve our future selves and past selves. It would not make sense, on this

[30] Korsgaard (2003)

view, to talk about two "selves" within one person differentiated by weakened psychological connections over time when our actions and projects necessarily involve identifying with future and former selves. When we choose to enter into relationships, careers, and other projects, we "presuppose and construct a continuity of identity and of agency."[31]

This concept of a person as centrally defined by his or her projects and actions privileges a special type of psychological connectedness that is left out in Parfit's picture. The relations that Korsgaard holds as crucial to understanding persons are ones that are authorial in nature. How does this connection differ from the psychological connections Parfit emphasizes? Parfit says that there is a trivial way in which a person at t1 and t2 remains the same: the person at the latter stage has the same beliefs as the person at the former stage. But beliefs can differ from one another in crucial ways. There are beliefs, for instance, that we merely acquired as children and which have persisted simply because they have remained unexamined. Then there are also the beliefs that persist because we choose them and endorse them. The latter kind of belief is more important to Korsgaard, who holds that authorship most truly defines the person who authors these beliefs.

Korsgaard rejects Parfit's claim that an account of personhood must be grounded in metaphysical facts in order to be justified. Korsgaard notes that

---

[31] Korsgaard (2003): 170.

when Parfit tries to argue from metaphysical facts to normative reasons, he ignores the fact that this in itself needs justification. Korsgaard argues differently, saying that there are other grounds for determining which actions can be properly ascribed to one, and that metaphysical facts do not constitute the sole ground for this determination. There is a practical ground that allows us to determine the nature of agency. On this argument, it is because we must carry out plans over time that we are properly called persons. This reverses the view that because we are persons, first and foremost, we have to carry out plans over time.

Korsgaard uses a group analogy to illustrate this point[32]: Suppose there are different agents that occupy my body. All of these agents must cooperate, so the unity of life is forced upon these agents in virtue of their shared embodiment. Because there is a single life that has to be led, these agents get together and cooperate, some making sacrifices for the benefit of others, and all of these compromises are reflected by a single self. This analogy illustrates Korsgaard's claim that the practical necessity of carrying one's plans and commitments into the future makes us agents.

Of course, one might object by arguing that we could simply call this conglomeration of agents a team rather than a person. Indeed, we seem to have distinct uses for both, and if a team is all that this is, we ought not to

[32] Korsgaard (2003): 171.

force this analogy onto our understanding of persons. But it is important to note that to make such an objection is to presume that the metaphysical facts override the essential embodiment these “agents” share. Korsgaard raises this analogy to bring out the argument that the necessity of making choices and seeing oneself as an agent does not depend upon underlying metaphysical facts. The fact that Korsgaard illustrates the possibility of an imaginary “team” of persons does not change her analogy: both persons, as we normally construe them, and this team-person, are necessarily agents because they face choices as a whole unit. The unity of life is forced upon each agent in this team-person equally, and because there is no escaping the necessity of this unity, each member has to work together and address choices together. And this forced embodiment, combined with the necessity to make decisions as a whole unit, is what makes this group essentially a single agent.

As Korsgaard notes, the reductionist argument implies that we can, in a sense, take inventory of our experiences and then, secondarily, ask what unifies them. But this way of talking about our experiences is misleading. Parfit assumes that the unity of consciousness requires a continuing psychological subject. This allows him the intuitive leverage he gains in the physics exam case, where he shows that when we separate the two hemispheres, we no longer have a unity of consciousness that can be appealed to by attributing this to a single subject.

Korsgaard argues that the only way Parfit gets away with this argument is by assuming that the only way to explain the unity of consciousness is to appeal to the metaphysical facts. And because Parfit shows that the metaphysical facts that might justify a unity of consciousness are absent, then there is no unity of consciousness to which one can appeal when making claims about persons. Korsgaard rejects Parfit's claim that the unity of consciousness is something that needs metaphysical explanation. She instead argues that the unity of consciousness is actually a *feature* of actions – such as perception, thinking, and acting – rather than an *enabler* of these activities. So consciousness, then, is a feature that emerges out of our various activities, "a feature of certain activities which percipient animals can perform."[33] And the performance of these very activities requires the unity of agency. This is because the unity of agency comes in part from the "raw necessity"[34] of our various motives and experiences being unified within a single body.

Since the two hemispheres of my brain share one body, it is necessary that they work together. So, Parfit argues, this unity is merely the result of the forced necessity of these two spheres working together. But, the unity of agency consists in something else that Parfit overlooks: that is, the unity inherent in deliberation. "To be sure, when I engage in psychic activities

[33] Korsgaard (2003): 174.
[34] Korsgaard (2003): 169.

deliberately, I regard myself as the subject of these activities."[35] In order to perform any activity, one must believe oneself to be the arbiter of these actions and choices.

The reductionist can argue that the only reason we believe that an action or choice has a subject is because of our language, and that the mere ability of our language to assign a subject to an activity is not anything important. But, Korsgaard argues, from a practical standpoint, actions and choices must be viewed as having subjects. And merely explaining this away by a theoretical approach does not detract from the importance of having an authorial stance. "It is only from the practical point of view that actions and choices can be distinguished from mere 'behavior' determined by biological and psychological laws. This does not mean that our existence as agents is asserted as a further fact, or requires a separately existing entity that should be discernable from the theoretical point of view. It is rather that from the practical point of view our relationship to our actions and choices is essentially *authorial*: from it, we view them as *our own*."[36] Korsgaard is suggesting here that the only intelligible way of understanding how actions are essentially authored is by assuming the authorial, or first-person, perspective.

---

[35] Korsgaard (2003): 175.
[36] Korsgaard (2003): 176-7.

Korsgaard ultimately argues that the authorial relation is crucial to understanding personal identity, but is entirely left out in the reductionist picture. Unlike the relations that comprise Relation R, Korsgaard argues, this authorial relation makes it necessary that there be a subject of actions, as it implies a communication between the person and its parts, and the person's ultimate authority over its parts.

An Assessment of Korsgaard's Strategy

Korsgaard's response to Parfit does not succeed in rebutting the coherence of Parfit's reductionist claims. This is a result of Korsgaard's account of the coexistence of the two perspectives we can take to persons. Her inability to rebut the reductionist account on its own terms results in a failure on Korsgaard's behalf to properly argue against Parfit's claims from the theoretical perspective.

Korsgaard insists that the coexistence of these two perspectives ought not to show that the two are contradictory when taken together. Indeed, she argues, the two perspectives "cannot be completely assimilated to each other, and the way we view ourselves when we occupy one can appear incongruous. The incongruity need not become contradiction, so long as we keep in mind that the two views of ourselves spring from two different relations in which

we stand to our actions."[37] Here she is quite willing to grant the independence of the two perspectives, and it is yet to be settled why one ought to be privileged over the other. This strategy undermines Korsgaard's emphasis on the practical perspective. Since she establishes the merits of each position, she grants the possibility of the theoretical perspective. In so doing, she grants Parfit the claim that *metaphysically* there are no persons, and so despite her insistence on the practical necessity of presupposing that there are persons for agency, she grants that that presupposition is metaphysically false and that the conception of agency she advances is, then, in some sense illusory.

Korsgaard cannot properly argue that the practical perspective is a better way to view persons *in general*, or vice versa, if the two are based on different relations we have to our actions and are thusly virtually incomparable. So far, all Korsgaard has done is undermine Parfit's implicit argument that the *only* perspective we can take toward persons is the theoretical perspective. But it isn't even clear that this is what Parfit is aiming for. Surely, Parfit can admit that we can take a kind of authorial attitude toward our actions, where choices seem immediate and necessarily demand resolution via our decisions. But all Parfit has to do is say that, when precisifying what matters about persons in thought experiments when personal identity is indeterminate, it is the theoretical perspective that will succeed *for*

[37] Korsgaard (2003): 176.

these purposes. This way, Parfit can suspend Korsgaard's arguments in favor of the practical perspective by just digging in his heels and insisting upon the reasonableness of the theoretical perspective for his purposes.

Korsgaard at different instances, however, seems to argue that the practical perspective is indeed the only perspective we can legitimately take toward persons. She does this by trying to extend her argument for the practical standpoint to explain moral reasoning. Korsgaard's practical standpoint is essentially an account of a personal relationship one can take toward his/her own actions. But even though she has insisted that the practical standpoint is based on certain unique relations an agent has to his/her actions, she hints that the idea of persons from any other perspective without reference to the practical standpoint might be incoherent. In *The Sources of Normativity*, Korsgaard makes a point about the necessity of the deliberative perspective for properly understanding normativity. But the analogy Korsgaard draws upon between her view of persons and reasons is clear throughout her work, and we can make inferences about how this view of normativity is structured to reflect her view of persons. In *Sources*, she argues, "Value, like freedom, is only directly accessible from within the standpoint of reflective consciousness . . . From [the] external, third-person perspective, all we can say is that when we are in the first-person perspective we find ourselves to be valuable, rather

than simply that we are valuable."[38] Here, Korsgaard clearly rejects the idea of objective normativity, and so, it would seem, the idea of an objective view of persons as having any real import in a discourse about personal identity. But Korsgaard never seems to mount an actual argument in support of this point. In order to show that an account of persons is most properly understood via the practical standpoint, she has to show that Parfit's theoretical standpoint is incoherent on its own terms. So far she has not shown that Parfit's theoretical perspective actually fails, nor has she argued against the possibility of the theoretical stance.

Even though Korsgaard's account of the practical perspective does not succeed at rebutting Parfit's claims, we can assess her account on its own terms. Perhaps if her account were more convincing, it would outweigh the plausibility of Parfit's account of reductionism. Korsgaard's claim that the practical perspective can be taken to apply to the larger realm of morality, however, does not seem plausible. She says, "Trying to see the value of humanity from the third-person perspective is like trying to see the colors someone sees by cracking open his skull. From outside, all we can say is why he sees them."[39] But this is clearly a dubious claim. We almost *exclusively* assess individuals from the third-person perspective when considering the reasonableness of a law. Any consideration of persons as collective members

[38] Korsgaard (1996): 124.
[39] Ibid.

of a whole, in order to assess laws and moral reasons, depends upon a picture that precisely guards against the essentially subjective view that Korsgaard advocates. Even Nagel, who is sympathetic to the value of the practical standpoint, argues that "giving the last word to the first person is a mistake."[40] Indeed, it seems obvious that there certainly are instances where a theoretical perspective is necessary in order to enact reasonable and applicable laws.

Korsgaard's account, then, gets us nowhere in rebutting Parfit's positive claims about the implications of reductionism for morality, nor does it amount to any useful account on its own terms. And in effect, the inability of Korsgaard's practical perspective argument to give us a picture of how we might view others as agents makes Parfit's conclusions about morality under the reductionist perspective stronger simply because Parfit's theoretical approach toward persons allows him a kind of objectivity that Korsgaard's account does not.

In Chapter Three, I will explicate Hume's account of personal identity and show how it both ties in with some of the strengths of the Kantian approach, but ultimately succeeds in amounting to a defensible rebuttal of Parfit's reductionist argument and his conclusions about agency and morality.

---

[40] Nagel (1996): 205.

Chapter Three: Hume's Fictionalism

Parfit's reductionist account of personal identity invites criticism because of his claim that our moral and prudential reasoning ought to reflect whatever metaphysical conclusions we draw regarding personal identity. As we saw in the previous chapter, neo-Kantians such as Korsgaard have resuscitated the Kantian argument that the reductionist position with respect to personal identity leaves a crucial metaphysical piece out of the puzzle, viz., the unity of agency. Korsgaard argues that Parfit's argument does not succeed because Parfit's metaphysical slicing and dicing omits this crucial ingredient.

Korsgaard and Parfit seem to agree on the presupposed premise that a purely metaphysical elucidation of the concept of personal identity is not only possible, but that it necessarily determines our theories about moral and prudential concern. They only disagree about the analysis, and so about the relevant consequences.

The neo-Kantian argument against Parfit, however, seems question-begging. Korsgaard's arguments in favor of a unity of agency seems promising because Korsgaard argues that we must examine our concepts of agency, as they are informed by and made intelligible through a network of social practices. But her project is limited because she bases her argument on the Kantian argument for the unity of consciousness, which is the very

argument Parfit takes himself to have rebutted, and she has no reply to his rebuttal.

In what follows, I will consider Hume's account of personal identity. Hume approaches the question of personal identity in two different ways. In the beginning of Book One of the *Treatise*, he claims to separate the question of personal identity into two: roughly, the theoretical question and the practical question. Book One, then, is devoted to the theoretical question, and Hume attacks this question by getting at the metaphysical underpinnings of personal identity. Both Hume and Parfit claim that there is a sense in which we can reduce persons to being mere aggregates of more particular psychological connections or experiences. For these reasons, many have held that Hume's account mirrors Parfit's reductionist picture of personal identity; indeed, Parfit himself has claimed a similarity between reductionism and Hume's account of personal identity in Book One.

I will argue, however, that Hume's approach to the theoretical question of personal identity is subtler than Parfit's. Hume develops a fictionalist account of personal identity, which holds that we impose notions of the self onto more distinct psychological particulars, asserting the "simple supposition of their continu'd existence."[41] Hume calls these "simple suppositions," "fictions." And even though these fictions are in some sense

[41] Hume (1978):,T 198.

artificial, fictions are nonetheless reasonable and, indeed, underpin important truths. Being fictions, they are not verifiable. But the fact that we cannot verify the fiction of personal identity does not rule out the fact that we can make true claims about personal identity. Parfit, on the other hand, reveals his eliminitavist aims when he claims that precisifying the "matters of fact" about the psychological particulars of persons ought to render less plausible the fiction of personal identity.

So there is a key distinction to note from the outset: While both Parfit and Hume recognize that no concept of ontologically independent persons has an unproblematic ontological justification, they differ in their accounts of whether this matters. The difference between Hume and Parfit's stances is a difference in their views regarding the existence conditions of persons over time. On Hume's fictionalist stance, the existence conditions for persons over time include conventions. Parfit, on the other hand, argues that because conventions arise after and with respect to Relation R, Relation R is the existence condition for persons over time.

Hume's fictionalist account supports his approach to personal identity in Book Two. Hume does not accept the notion that a metaphysical picture of personal identity ought to motivate or change our views of personal identity as they exist in moral and prudential reasoning. In fact, Hume argues in precisely the opposite direction: that our conventions regarding morality and prudence

vindicate the conventions that constitute our identity. Ironically, Hume will turn out to be more successful than the neo-Kantians in arguing from a priority of the practical over the theoretical in favor of a robust concept of personal identity, and will turn out to be farther from Parfit's apparently Humean view than are the neo-Kantians.

Hume's Book One: A "Bundle" Theory of Persons

Hume's approach to personal identity appears at first to be eliminative, as does Parfit's own account. Unlike Parfit, however, Hume argues that showing the concept of personal identity to be nothing more than a fiction does not entail that our practices involving and approaches toward personal identity must be changed. When Hume argues that personal identity is a fiction constituted by other facts, he is not thereby arguing that personal identity does not play its own important fact-constituting role. Indeed, the fiction of personal identity determines certain truths that are irreducible to the truths of its constitutive facts. In what follows, I will examine Hume's view of personal identity as it develops in Book One. I will then explicate Hume's notion of fictionalism, and show how his view of persons can be understood with respect to this model.

Hume begins his discussion of persons by asking how it is that we get the impression of a self. His approach toward the nature of personal identity is a novel one analogous to his discussion of the external world. A proper

skeptical approach to understanding the nature of our beliefs about the external world involves elucidating the ways in which we have arrived at these beliefs. It would be fruitless to pursue this question by asking whether or not an external world exists at all. We cannot entirely throw our beliefs about the external world into doubt to the extent that we can act as if, upon reflection, it is a discoverable fact whether or not the world exists. To question our beliefs about the external world by asserting that its very existence is dubious is conceptually impossible, simply because its existence grounds all natural human beliefs and practices, including reasoning. So, as Hume argues, we cannot help but grant the basic premise that the world exists. Hume says, with regard to our beliefs about the external world, "We may well ask, *What causes induce us to believe in the existence of body?* But 'tis in vain to ask, *Whether there be body or not?* That is a point, which we must take for granted in all our reasonings."[42]

Hume's approach to the skeptical inquiries into the existence of the external world is analogous to his approach to questions about personal identity. Some skeptical approaches to this question have begun by questioning whether there even exists such a thing as a person, or an "I." Like the analogous skeptical questions about the existence of the external world, to ask whether or not persons exist at all is a fruitless question.

---

[42] Hume (1978): T 187.

In his discussion of the external world, Hume distinguishes the questions, “How have I arrived at this belief in P?” and “Does P exist?” This distinction can be carried over to apply to our notion of persons. We can come to understand personal identity by looking at what causes us to have the concept. But it is indeed an absurd question to ask whether or not persons as we normally conceive of them exist, because clearly we employ the notion of persons in our reasoning, and have thus granted these notions some kind of status in order to make sense of the world. Understanding the distinction Hume makes between proper and improper skeptical questions regarding personal identity will help guide us through Hume’s substantive discussion of personal identity and, later, his fictionalist stance.

First, Hume asks how it is that we have an impression of the self. Hume argues that when we examine the origin of our impression of the self, we run into contradictions. If any one impression is said to be the cause of our notion of personal identity, this impression must “continue invariably the same, through the whole course of our lives; since self is supposed to exist after that manner.”[43] But we see that there is no one impression that persists unchanged so as to allow for the kind of constancy our notion of the self requires. Pointing to notions like the passions and sensations to explain the impression of a self will fail, since these are all ephemeral. As such, Hume

[43] Hume (1978): T 162.

questions how they could be connected with the idea of the self. He argues that when we reflect on what we call ourselves, we inevitably start talking about these passing perceptions, feelings, thoughts, and so on. And indeed, he says, we can never have a notion of the self without making reference to some kind of thought or sensory experience. So, Hume argues that persons are nothing "but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement."[44] We cannot find, beyond these separate, temporary impressions, any strict notion of the self that does not make reference to our experiences. Here Hume is clearly rejecting the claim there is any kind of deep metaphysical entity that underlies or supports these disparate impressions.

Hume then tries to reconcile the idea that persons are nothing but bundles of perceptions with the traditional notion of personal identity – that a person must be a single object over time. Here is where the contradiction lies: While it is clear that upon rational reflection, we can distinguish all of our various successive impressions, we still have a propensity to conjoin them and attribute a kind of relation to these that reflects constancy, a fundamental part of our notion of the self.

This contradiction is readily identifiable in our concept of simple objects. Hume argues that when considering the identity of objects over time,

---

[44] Ibid.

we often fail to distinguish between numeric and qualitative identity. Numeric identity is the identity of any object to only one object, viz., itself, even if this object changes over time. Qualitative identity, on the other hand, can obtain between two objects that, while not one in the same, share exactly the same qualities (for example, two billiard balls that are identical in color, shape, and so on, but are numerically distinct). So while one might hear numerically distinct sounds, all alike in quality, one might be apt to ascribe these different sounds to one origin. Hume identifies a number of factors that determine the degree to which we confer identity on a single object over time. An object A at time T1 would be numerically identical to itself at T2 if, and only if, all of the physical properties of A at T1 are identical to those it has at T2. So consider an object assembled of various smaller parts. If we remove or alter any one of these constitutive parts, we have ruined the identity over time of the object. Indeed, even if the object *ages,* its later stage fails to be numerically identical with its earlier stage, demonstrating that numerical identity over time is chimerical. But clearly we continue to ascribe identity over time to objects that change. Hume argues that the rate and proportion of change play a decisive role in our reasoning about identity of objects over time.

While these two factors ought not make a difference with respect to the numerical identity of an object over time, Hume is arguing that these

factors nevertheless guide *our propensity to ascribe* identity. On the other hand, Hume notes, if an object were to undergo a substantial change, such as the addition of large appendage, this radical change might factor into our ability to ascribe identity to the object over time. Moreover, it seems as though changes that occur gradually make us more apt to continue ascribing identity to an object. For example, consider the classic ship of Theseus puzzle. The ship was originally built of certain planks of wood, but over time, these planks were gradually replaced, one by one. The gradual nature of this change seems to allow us to still claim that the ship now and the ship 30 years ago are one and the same, even if the original ship has been entirely replaced. Ultimately, the mind somehow sees a clear passage from one instance of an object to a later instance and does not see these two instances to be distinct. It is from this connection that we arrive at the notion of identity. The ship of Theseus example shows that although an object might undergo radical change, we are still in some circumstances prone to ascribe numeric identity to a single object, confusing a legitimate ascription of qualitative identity with the discovery of a nonexistent numerical identity. A similar confusion underlies the supposition that there is a numerically identical self that exists over time over and above the qualitative identity between our stages.

Hume then compares the identity of objects to personal identity. Objects, like persons, are things that we assume to be constant and invariable

over time. This assumption is what Hume calls the notion of identity or sameness. At the same time, however, we understand the idea that over time, an object can change, so much so that we can demarcate different objects existing in succession. This is Hume's notion of diversity. The relation between these two contrary assumptions, as they work together to build our notion of identity, is explicit when we consider Leibniz's Law. According to Leibniz's Law, an object can only be numerically identical to itself, since things are identical to one another if, and only if, they are indiscernible. But, as an object changes over time, we cannot say that the successive stages of an object undergoing change are numerically identical to each other, since by definition they are discernible. Indeed, this would collapse the notion of change over time. Instead, while these successive changes are numerically diverse, they can be identical in kind. That is, they share some characteristics salient to us in terms of which we regard them as similar.

When we say that an object remains the same over time, we allow that the successive stages of change in an object be numerically distinct as long as we can preserve the notion of a kind of qualitative identity between these successive stages. So, a tree, for example, can change over time. These stages of change are numerically distinct, but we unite these successive stages in virtue of the fact that they share many of the same qualities (such as being a poplar, being descendants of a seed planted on a particular date in a particular

place, etc). In the act of attributing identity to what are, strictly speaking, very distinct impressions, we are bringing together the notion of diversity and identity. Hume argues that it is absurd that we recognize, yet conflate, the notions of diversity and identity by the same faculty of imagination. In order to do this, Hume maintains that we come up with a kind of fictional idea of identity. The problem of how it is that we assign identity to objects in spite of whatever radical changes they incur over time is coupled with the problem regarding how it is that we assume objects to persist even when our attention is directed elsewhere. When we perceive an object, and then divert our attention when the content of our perceptual field changes, we assume that the object we perceived first has independent existence and hence continues "uninterrupted."[45] But of course, there is no sense in which we can verify that an object persists uninterrupted when we are asleep or not present. How is it, then, that we assign identity to a single object when our perceptions of it may be interrupted and, thus, are numerically distinct?

Hume argues that we make two assumptions: First, when we perceive change, we regard change as happening to a single object over time. Second, when our attention to an object is interrupted, the existence of the object itself is not therefore interrupted, but rather persists independently. In order to deal with these assumptions, in light of the fact that these assumptions cannot be

[45] Hume (1978): T 254.

verified, Hume says that "in order to justify to ourselves this absurdity, we often feign some new and unintelligible principle, that connects the objects together, and prevents their interruption or variation. Thus, we feign the continued existence of the perceptions of our senses to remove the interruption; and run into the notion of a *soul*, and *self*, and *substance*, to disguise the variation."[46]

These "feigned" notions are what Hume calls fictions. But it is important to note that by "fiction," Hume does not mean that these notions are false; rather, they are simply unverifiable. Of course, there is no way to assure that objects persist independently, nor is it possible to show that there are simple, unchanging objects to which we can attribute change. Yet, as Annette Baier argues, "the postulates of the independence of the world from our observations, and of the background presence of something that is invariant in all our mind's variations, seem to make factual claims."[47] Indeed, fictions can be thought of as "plausible stories we tell ourselves" in order to reconcile paradoxical claims about our environment and persistence. Fictions are useful constructs that enable us to speak of things like enduring objects and the external world in light of whatever conflicting claims we simultaneously entertain when we try to illuminate these ideas. Fictions, then, are grounded by a constitutive base of facts, but are irreducible to these facts. Indeed,

---

[46] Ibid.
[47] Baier (1991): 103.

beginning with various perceptions and relations of perceptions, we form a kind of "'system' that goes beyond that we strictly know to be true."[48] This system is what ultimately yields the fictional components of our beliefs. But fictions cannot be reduced to these perceptions, precisely because fictions accomplish extra conceptual work that relations of perceptions by themselves cannot provide.

To see how the fiction of identity accomplishes this extra conceptual work, consider, for example, the relation of resemblance. We perceive this relation when we consider, for example, a tree persisting over time. We note that Tree 1 at Time 1 (T1) has a certain set of properties that is qualitatively identical to the set of properties of Tree 2 at T2. But suppose that we have left the location of the tree in between T1 and T2. Hence, when we say we are perceiving a tree at T1 and T2, these perceptions are numerically distinct but qualitatively identical. From the relation of resemblance, we feign the fiction of a single tree that persists in virtue of the qualitative identity we have discerned.

To get an even clearer picture of how fictions work, consider the relationship between fictional characters in novels and whatever facts give rise to them. Fictional characters in novels are not reducible to their constitutive non-fictional phenomena. So while Holden Caulfield is entirely

[48] Baier (1991): 104.

constituted by J.D. Salinger and the more particular facts about text of the *The Catcher in the Rye*, Holden Caulfield as a fiction is neither reducible to facts about the book, nor Salinger's relevant mental states, attitudes, and ideas at the time of writing *The Catcher in the Rye*. Indeed, as a fiction, the text has its own specific fact-constituting potential. We can make true claims about Holden Caulfield: He is a disgruntled teenager; he dropped out of his private high school, etc. But the truth or falsity of these claims do not depend upon facts about J.D. Salinger.

This fiction cannot be reduced to the relation of resemblance we perceived, because the fiction (namely, that there is one tree persisting over time) has, in virtue of its fictional nature, its own fact-constituting potential. Once we have granted the fiction of the single tree, we can make true claims about this tree ("this Tree is a poplar," "this Tree is 50 years old") that we could not make simply by appealing to the relation of resemblance. The relation of resemblance, in this example, only tells us that there are certain properties that are qualitatively identical between a tree at T1 and a tree at T2, but does not assert that these "two trees" are one and the same. In this way, the fiction of the single persisting tree is not reducible to the relation of resemblance, despite being constituted by this relation. The fiction is reinforced by conventions that reflect the constitutive relations of a fiction. When we perceive the relation of resemblance in this case, we develop certain

conventions, such as the naming of this tree, specifying the tree's exact location, etc. All of these claims reflect the relation of resemblance, and in using these conventions repeatedly, we arrive at the fiction of the single tree persisting over time.

Similarly, an analysis of persons may reveal that nothing over and above certain psychological relations and conventions constitute persons, but the facts about this constitutive base do not make it such that the supervening person whose identity is so constituted can be reduced to the constituents of the supervenience base. Indeed, we can state truths about persons and not commit ourselves to the truth or falsity of statements about that upon which they supervene. The fiction of personal identity, then, possesses a specific fact-constituting potential irreducible to the facts about constitutive relations or conventions. Additionally, the fiction of personal identity has a kind of practical reasonableness. This point will be further elucidated in the later discussion of Hume's passions. After reflecting upon the paradoxical notions that comprise our concept of identity, Hume turns toward the problem of personal identity over time. He says that our notion of personal identity takes much the same shape as our notions of identity of objects over time: "The identity, which we ascribe to the mind of man, is only a fictitious one, and of a like kind with that which we ascribe to vegetables and animal bodies. It . . .

must proceed from a like operation of the imagination upon like objects."[49] As with objects, the mind fails to notice a series of small changes that occur in a person over time, and because this transition between these changes is an easy one, the mind believes that it is dealing with a single person.

Consider, for example, the idea that we consider Jones at T1 to be the same Jones at T2, despite the fact that in between these two times, Jones has undergone some physical change. But even though we believe we are dealing with one and the same person, the mind is also capable of viewing a person from two separate points of view once a significant change has occurred. For example, to play the hard-line skeptic, we can in fact note what physical changes are taking place in Jones, and, as a result, declare that identity between the two Joneses at these two stages does not obtain.

To prevent us from making the latter claim (or from remaining in a perpetual state of painful ambivalence about which stance to take), the mind forms the fiction of personal identity. For Hume, the notion of personal identity is a fiction constituted by relations and conventions. The mind cannot bring different impressions together and make these impressions lose their distinctiveness – as in the case where the mind is able to perceive of changes and hence demarcate different objects at different times. But in spite of the uniqueness of every impression we have, when we still presuppose "the whole

[49] Hume (1978): T 259.

train of perceptions to be united by identity, a question naturally arises concerning this relation of identity, whether it be something that really binds our several perceptions together, or only associates their ideas in the imagination; that is, in other words, whether, in pronouncing concerning the identity of a person, we observe some real bond among his perceptions, or only feel one among the ideas we form of them."[50] Hume argues that personal identity over time depends on certain relations, including resemblance, contiguity and causation.[51] These relations form the basis of the ascription of qualitative identity to an object. These relations then give rise to a development of conventions. An example of a convention about persons is that we regard future stages of ourselves as identical in kind to our present stage. Causal relations cause us to believe ourselves to be temporally extended persons. Without causal relations, the view that we have any future continuers, or any persisting self at all, becomes impossible, because causal relations allow for us to take successively existing objects to be single objects that endure over time in spite of change.

Similarly, the relation of resemblance supports this convention, because it enables us to make qualitative comparisons between two stages of a person and unite them into one. So, conventions, such as the one described, are constituted by relations, and in turn, these conventions give rise to the

[50] Hume (1978): T 168.
[51] Hume (1978): T 260.

fiction of personal identity: namely, that there is a single person persisting over time to whom we can ascribe change. The fiction of personal identity is reinforced by the beliefs that arise from our conventions. For example, as a result of the convention that persons regard their future stages as qualitatively identical, persons develop attitudes of self-concern toward these continuants. These beliefs help reinforce the fiction of personal identity. Most importantly, the pervasiveness of conventions about persons and the beliefs that result from and reinforce these conventions make the fiction of personal identity natural and relevant.

These relations, coupled with those conventions that have arisen in dependence on these relations, give rise to the fiction of the self. The self, then, is grounded in facts about our relations and conventions, but although the self we construct supervenes on these facts, the self is not reducible to these constituents. We developed this idea earlier when discussing the specific fact-constituting discourse that evolves when we posit the fiction of the single existing tree over time. The structure of Hume's fictionalist account of personal identity is clearest in Hume's analogy between personal identity and nationhood. As we have seen, this is an analogy that Parfit takes from Hume to defend his reductionism, suggesting that their two accounts of personal identity are similar. But their accounts differ in crucial ways. From the fictionalist standpoint, nations are fictions that just exist when territories,

persons, and conventions exist. A nation may only exist just when certain persons act in certain ways, its territory boundaries are mapped out in certain ways, and so on. When all of the relevant particulars that underlie the construct of a nation obtain, we can say that a nation exists. If we were able to get rid of the particulars that ground a nation, then the nation might cease to exist. But this does not mean that nations are reducible to these constitutive facts. We can make claims about nations that cannot be made simply by appealing to every fact that constitutes a nation.

This goes for persons as well. Were we to abolish the relevant conventions that constitute persons, we would no longer have persons. Persons are logically dependent upon conventions, but, just like nations, they are real, in virtue of the fact that the conventions and relations that underlie their constitution are real. Additionally, claims can be made about persons that are irreducible to all of the further facts that constitute persons. Parfit, on the other hand, holds that because a nation is a supervening fact that just consists in further facts, these further facts are *all that matters*. A nation, for Parfit, is constituted by more particular facts regarding its citizens, boundaries, laws, etc. But "a nation is not an entity that exists separately, apart from its citizens and its territory."[52] Similarly, persons are constituted by more particular psychological and physical facts, but persons are not

[52] Parfit (1984): 211.

separately existing entities. As with nations, Parfit argues that persons could be fully explained by an elucidation of these more particular facts. Essentially, facts about persons are *reducible to* facts about their constitutive base, because persons fail to be separately existing entities.

Parfit's argument here about what matters seems to go as follows:

(P1) Personal identity is a supervening fact that consists in more particular psychological and physical facts.

(P2) The only *genuine* facts are those that are metaphysically basic.

(P3) The psychological and physical facts that constitute personal identity are metaphysically basic. *So,*

(P4) Personal identity does not *in itself* have metaphysical reality apart from that of the psychological and physical facts upon which it supervenes.

(C) Therefore, only the psychological and physical constitutive facts are metaphysically real.

On the basis of this conclusion, Parfit argues that a complete description of reality could omit persons, because the truths about persons are reducible to the truths about its constitutive base.[53]

The difference between the Humean and Parfitian stances toward the nation analogy should by now be clear. Hume's fictionalism bases the fiction of personal identity in relations and conventions, but does not deny that true claims can be made about persons that are irreducible to its constitutive base.

[53] Parfit (1984): 212.

Hume would therefore reject (P4) as a result of denying (P2), which he denies by emphasizing the extent to which fictionalism offers us an account of reality that Parfit takes for granted. So, by rejecting two of these premises, Hume undermines Parfit's conclusion. The mere fact that more particular psychological and physical relations and conventions constitute persons does not imply that we could give up our talk of persons in favor of more particular talk of relations and conventions.

Hume concludes Book One with the claim that personal identity is a fiction constituted by further facts. Because relations allow for easy transitions from one impression to another impression, we assign identity to these impressions. But since these relations can change or come in degrees, we cannot pick a time at which or a degree to which the relation loses or gains its title of being constitutive of personal identity. So, Hume concludes, metaphysical questions concerning personal identity "are merely verbal, except so far as the relation of parts gives rise to some fiction or imaginary principle of union, as we have already observed."[54] When we have the fiction of personal identity, then, we are in an entirely different fact-constituting discourse than the discourse in which we may talk about the more particular relations underlying personal identity. Hume concludes that although our concept of the self is not that of an independent metaphysical entity, it is that

[54] Hume (1978): T 262.

of an indispensable fiction that drives our lives and grounds important social practices and truths. We cannot understand human life without it. Hume hence argues that we do not need to assign personal identity a metaphysical status in order to regard it as genuine. Indeed, looking for a single referent that grounds the notion of personal identity is absurd: “No connexions among distinct existences are ever discoverable by human understanding. We only feel a connexion or determination of the thought, to pass from one object to another. It follows, therefore, that the thought alone finds personal identity, when reflecting on the train of past perceptions that compose a mind, the ideas of them are felt to be connected together, and naturally introduce each other.”[55]

So, for Hume, relations give rise to certain conventions, which in turn give rise to the fiction of personal identity. But while personal identity is brought about and hence constituted by these further particulars, it is not the case that the truths of personal identity are reducible to the truths of its constitutive base. Fiction constitutes its own realm of discourse and there are true claims that can be made about it that are not reducible to its constitutive facts.

Hume’s Book Two: Persons and their Passions

While Hume treated the issue of personal identity in Book One from a theoretical point of view, Book Two of the *Treatise* approaches the problem

[55] Hume (1978): T 635.

of personal identity from a practical standpoint. Indeed, Hume notes that the subject of Book One will be personal identity "as it regards our thought and imagination," and, later, in Book Two, "as it regards our passions or the concern we take in ourselves."[56] Hume's focus in Book Two is how the self is "actuated" by passions, and how it is central to these passions that we interact with other selves and treat these other selves as temporally extended, unified persons. The practical utility of the concept of personhood, in terms of prudential and moral reasoning, is supported by the fictionalist account Hume develops in Book One. Hume defines passions as reflective impressions that arise as a result of an original impression, which would be a bodily sensation or any kind of general sensation that arises without an antecedent. Hume then differentiates between the cause of a passion and its object.[57] One's qualities, like judgment, memory, and disposition, can be among these causes. For example, possessing excellent memory can be a cause of pride. Similarly, a cause can come in the form of an object – using the example of pride, I can be proud of an object I possess, and this object can be properly called the cause of this passion.

Passions then fix upon an object – the self. It is here that Hume reveals that having the self as the object of the passions is wholly natural, as they constitute and reflect the idea of the self. The passions only become

[56] Hume (1978): T 253.
[57] Hume (1978): T 278.

intelligible within a network of individuals, whereby one realizes oneself as a person among persons. This is illuminated most clearly when Hume discusses how the passions of pride and humility, and the passions love and hatred, are brought together by their focus on persons: “As the immediate object of pride and humility is self or that identical person, of whose thoughts, actions, and sensations we are intimately conscious; so the object of love and hatred is some other person, of whose thoughts, actions, and sensations we are not conscious.”[58] Here we see that all passions fix upon objects and can only be rendered intelligible when they have the self, or other selves, as their referents. The distinction in perspective between seeing persons as bundles of perceptions and seeing persons as actuated by their passions is significant, but the two discussions are not inconsistent. There is a legitimate sense in which persons are merely collections of closely related but still distinct sensory phenomena and impressions, but there is also a sense in which, as a consequence of convention, our concept of the self as a steady referent and object plays an important role, since it serves as the object of our passions. It would be absurd to try to understand the passions without referring to the self, to which the passions always refer. Baier emphasizes this point: “If reason is and ought to be the slave of the passions, it is not going to be able to get an adequate idea of the self, one of whose ‘organs’ it is, if it tries to abstract from the passions, those more vital and more dominant organs of the mind and

[58] Hume (1978): T 329.

person."[59] Here we see that it is central to Hume's discussion of the passions that the idea of the self is taken seriously, as it is the basis of our own desires and allows us to conceive ourselves as participating in a community of other persons with whom we interact.

Baier points out that Hume regards the body as the most proper picture of the self. This allows us to understand how the qualities of our minds and bodies (as mentioned before, things like memory and disposition) can be proper causes of the passions. It also allows for a clearer picture of how the self can be a proper object of the passions. The passions must focus upon the self or selves, and the human body comes forward as the most natural and eligible referent for the passions. In this way, it is easier to understand how other selves can be the objects of the passions as well, because we naturally identify other persons with their bodies.

To make this point clearer: Imagine that you are filled with contempt for another person. Because your contempt needs a tangible, fixed object, you summon this person's body to mind, and this picture provides the simplest way to treat persons as the objects of our passions. So Hume is, in a way, appealing to the view that we simply are our bodies, since our ability to identify persons with their bodies provides us ground to make other persons the objects of our passions. Furthermore, we could not understand the self

[59] Baier (1991): 130.

without making reference not only to other persons, but also those objects and qualities that are subsumed as one's proper possessions. There is a special sense in which my particular personality traits, my love of certain food, my relationships with my friends, etc., are distinctly mine. Hume argues that there is no way to justify one's possession of these objects and qualities – rather, we have a simple, natural claim to them, and without them, the picture of the self loses all sense. Hume says, "Our self, independent of the perception of every other object, is in reality nothing: For which reason we must turn our view to external objects; and 'tis natural for us to consider with most attention such as lie contiguous to us, or resemble us."[60] Here Hume emphasizes the important extent to which we are surrounded by a world of objects and the way in which we come to understand the separateness of persons. The importance of objects and other persons lies in how they are related to us, and how we come to understand these relations as being not only specific to ourselves, but mirrored in the private lives of others. There is a way in which a person arrives at an understanding of what it means to possess an object, to have familial relations, to have loves, etc., by seeing also how other people have these relations themselves. As Baier says, it is because we see other persons as beings like ourselves, equipped with passions and who stand in special relations to others as we do, we formulate the concept of the self. And this understanding is aided by our reliance on bodies as being the natural referents of persons:

[60] Hume (1978): T 340-1.

"Since our conception of a fellow is of a flesh-and-blood person, then whatever conception we have of the identity over time of a living expressive body will be the core of our notion of a person's identity, and so of our own identities. Since we can see the separateness of human bodies . . . we know what makes one person different from another, and the experience of disagreement and conflict will reinforce that knowledge." [61]

Our ability to see persons as bodies is dependent upon a kind of network of selves to render the passions and our own relations to people intelligible. In a sense, we are always taking cues from others' lives and are engaged in an implicit dialogue with the actions and analogous relations of other persons, and this dialogue, to an important extent, informs us about our concept of personhood. Hume's account of personal identity as it is cashed out in his discussion of the passions is supported by his fictionalist arguments in Book One. In Book One, Hume grants that the notion of personal identity is a fiction constituted by relations of ideas and the conventions that result from and reflect these relations. Hume moves from explaining how the fiction of personal identity is constituted by relations and conventions, to demonstrating why this fiction is necessary.

We can glean three important insights from Hume's work: First, while personal identity is indeed a fiction, claims about persons are not therefore

---

[61] Baier (1991): 136.

false; on the contrary, many of them are reasonable and the fiction itself constitutes their truth conditions. Second, because it is a fiction, personhood is not reducible to more particular constitutive facts. Third, the concept of personal identity is necessary to and hence guides our understanding of human passions and hence of moral life. Without an account of persons (which Parfit claims is superfluous), the natural phenomena of passions would be entirely inexplicable.

Hume's account of personal identity is more plausible than Parfit's reductionism, because while it acknowledges, as Parfit does, the fact that personal identity is not defensibly secured by some kind of metaphysical appendage, it does not thereby rule out the possibility of our concepts of persons as having a legitimate role in a moral and prudential sense. Indeed, the problem with Parfit's argument lies precisely in his confusing account of what facts matter, and how they matter. The crucial difficulty with Parfit's argument is (P2): The only *genuine* facts are those that are metaphysically basic. As we saw in the earlier discussion of fictions, it is possible for there to be facts about fictions that are not metaphysically basic. Parfit has yet to prove how or why it is that only metaphysical facts are genuine facts.

Rejecting (P2) suggests a different way to proceed. On the one hand, we can conclude that personal identity exists *because* it is constituted by more particular psychological and physical facts. Indeed, the construct is real

because these constitutive facts are real. But, Parfit also holds that the construct of personal identity is *just reducible* to further facts. So, any statement about personal identity is *just reducible* to a statement about further facts. But we can now see the tension in Parfit's account: he simultaneously grants reality to personal identity in virtue of the reality of its constitutive facts, and deflates the reality of identity by holding that personal identity is reducible to and hence *nothing more than* these facts.

The implausibility of Parfit's argument, then, lies in the fact that what allows him to grant that persons are things that are reducible to their constitutive base is the premise that persons are real in virtue of the reality of their metaphysically basic constitutive facts. But if persons are real in any sense, then they cannot cease to be real just in virtue of being reducible to the very facts that endow them with reality. Hume, on the other hand, is safe from this difficult tension, because his fictionalist stance toward personal identity grants that persons, while fictions constituted by further facts, do have independent reality that make them irreducible to the very facts that, in virtue of their reality, constitute persons.

We can now see that Hume's account succeeds where Korsgaard's response appears question begging. Korsgaard's and Hume's accounts of personal identity both focus on the practical basis of our concept of persons. Korsgaard argues that because of the practical necessity of agency, we cannot

dispense with the concept of persons. But she locates the practical necessity of agency in the Kantian notion of the unity of consciousness, a notion that implies that the concept of persons must be grounded in metaphysical necessity. Korsgaard then falls prey to Parfit's arguments in arguing that personal identity is a metaphysical necessity, because all Parfit has to do is provide a counterexample, like the fission case, where we can undermine any appeal to metaphysics for an explanation of personal identity. Hume, on the other hand, merely grants that the notion of personal identity, while a fiction, is reasonable on its own terms. This allows Hume to press on Parfit's inability to show that the construct of personal identity is not independently reasonable.

Hume's Second Thoughts

As I have argued, Hume's development of personal identity in Book Two is easily reconcilable with his theoretical approach to personhood in Book One. Why, then, does Hume claim, in the Appendix, to be involved "in such a labyrinth," wherein he claims not to be able to render his opinions "consistent"?[62] Baier suggests that Hume only finds himself in this puzzle when he confines his questions about personal identity to Hume's "solipsistic intellectualist views"[63] in Book One. She argues that if we take the issue of personal identity to be an issue solely decided with respect to the kind of conclusions we drew about the theoretical nature of persons in Book One, we

find ourselves unable to arrive at any clear answer about the nature of persons. Put simply, it is not clear to Hume whether or not persons should be treated as the metaphysical "bundles" that we saw in Book One, or if they should be the persons involved in a complex network of interpersonal relations, as in Book Two.

There is an important sense in which our ordinary lives are guided by our simple reliance on bodies being persons. When I call out to you, I am relying on your body to tell me whom to pick out of a group of people. I do not rely on any complex metaphysical account of who you are and what differentiates you from others. This simple reliance on bodies as persons is what determines personal identity. In more complex situations, when it becomes a question of who is the true claimant of an identity, we can resort to the legal system to decide this matter. But, Baier argues, "these grammatical puzzles about identity are mostly intellectual, not real, dilemmas."[64] It is of no surprise that we find trouble applying our former concepts of persons in cases, such as teletransportation or brain-switching, because what we do in these cases is strip persons of their complex social roles and treat personal identity as if it were a purely metaphysical question. We hence abstract from the very features that make them persons, and have nothing left about which to reason.

---

[62] Hume (1978): T 633.
[63] Baier (1991): 138.
[64] Baier (1991): 138.

It is, then, Hume's focus on persons as, firstly, bodies, and consequently, as the objects of our passions, that Baier believes "banishes the ghost of that Book One worry" about what constitutes the self. The Appendix, she argues, is seeking and fails to create a sensible picture of persons by relying on metaphysics, but it is absurd to abstract persons away from other persons, in common social arenas and in the context of special interpersonal relations. This much said, it is important to reiterate that the conclusions Hume draws in Book Two do not annihilate the conclusions of Book One. Rather, Hume just makes it necessary that the persons we distinguished as bundles of impressions also be "inseparable from the system which is the living human body."[65] The purpose of Book Two is to make sense of what conclusions we drew in Book One by providing them a larger social context.

Conclusion

Parfit introduces his reductionist account of personal identity by implying it is Humean in spirit. I have argued instead that Parfit and Hume disagree sharply about both the nature of persons as constructions, and what facts are morally relevant. While Parfit and Hume agree that there is no ontological basis that can explain personal identity, they present radically different accounts about how this fact ought to affect our moral reasoning. For Parfit, in the absence of a deep metaphysical fact about personal identity,

[65] Baier (1991): 142.

personal identity is rendered useless. We ought, therefore, to revise our moral reasoning to reflect the metaphysical facts that constitute personal identity. Hume's fictionalist stance holds that persons are fictions, but that the fiction of personal identity is reasonable, even if it is not verifiable.

This chapter has served as a sketch and defense of Hume's fictionalism. In Chapter Four, I will show how Hume's fictionalist account of personal identity comes into play when we examine Parfit's revisionary moral outlook. Parfit's reductionist claims about how to restructure our laws and moral issues have found substantial opposition from defenders of morality as it is structured by personal identity. I will argue that the debates between these defenders of common sense morality and Parfit echo the essential debate between Parfit and Hume. Finally, I will show what resources Hume provides in evaluating Parfit's revisionary moral outlook.

Chapter Four: Reductionism and Morality

Parfit has argued that our traditional notions of personal identity are mistaken. Put simply, we cannot locate personal identity in any kind of simple metaphysical entity, like a Cartesian ego, and use this to give us decisive answers in problem cases involving personal identity. Parfit argues that persons instead just consist in more basic, impersonal psychological and physical facts. If personal identity just consists in these further facts, personal identity as such cannot fundamentally matter. Instead, what ought to matter to us are the psychological connections that obtain between different parts of a life, and psychological continuity over the span of a life.

In Chapters Two and Three, we considered Kantian and Humean objections to the implication that accepting reductionism ought to make us revise our moral and prudential beliefs. In Chapter Two, we saw that reductionism is seriously at odds with common sense egoism, or our traditional account of prudential rationality. On Parfit's view, once we pay attention to the psychological connections that obtain between different temporal parts, and emphasize the degree to which these psychological connections hold, it is irrational to consider a self existing at a temporal stage distantly related to one's current self as identical to one's present self.

As we saw applied to prudential reasoning in Chapter Two, Parfit argues that a reductionist account of personal identity can adopt either the Moderate Claim or the Extreme Claim. The Moderate Claim about morality states since Relation R is all that matters, we ought to demarcate agents in terms of how psychologically connected they are to each other. The Extreme Claim, on the other hand, states that once we accept reductionism, there is no reason to punish someone for an action they committed earlier. For reductionists who adopt the Extreme Claim, desert requires the deep further fact about identity. An understanding of each of these two alternatives, as well as Parfit's sympathy for the Moderate Claim about morality in *Reasons and Persons*, will structure this chapter.

The objections to Parfit that we examined in previous chapters sideline Parfit's arguments concerning the implications of reductionism because they undermine the premises that lead Parfit to reductionism. In this chapter we will grant Parfitian reductionism for the sake of argument, and examine its consequences in the moral sphere.

Reductionism has major implications for moral reasoning. Parfit argues that because the fact of personal identity is metaphysically "less deep" – since the unity of a life is simply the holding of various relations between experiences in a life – we ought to become concerned more about "the quality

of experiences, and less concerned about whose experiences they are."[66] Parfit hence argues that our notions of morality should change to be "more impersonal."[67] The tacit premise in Parfit's argument for an impersonal view of morality is that our beliefs about the nature of personal identity are the only rational basis of our moral beliefs. But because our beliefs about the nature of personal identity are mistaken, a revisionary account of morality, as well as of prudential concern, is in order.

In Chapter Two, we identified two premises in the argument for common sense egoism against which Parfit argues in his discussion of the effects of reductionism on rationality. These were:

> **(RP1)** A person has reason to make sure that the actions he performs benefit him, in virtue of the fact that these actions are his.
>
> **(RP2)** A rational agent has reason to be concerned about his future selves, in virtue of the fact that his future selves are temporal stages of one and the same person.

These premises find their parallels in the moral sphere. In common sense morality, there are two premises that Parfit wants to reject:

> **(MP1)** The separateness of persons matters morally.
>
> **(MP2)** Persons should be regarded as single, unified agents over time.

---

[66] Parfit (1984): 346.
[67] Parfit (1984): 443.

These two premises are central to a common sense view of morality, and are reflected in and reinforced by our legal practices. Of course, it is in keeping with our intuitions that it indeed matters to whom a benefit or a burden is assigned. Similarly, we tend to hold a person accountable for his/her crimes or promises across the span of his/her life, and attribute one's actions at different temporal stages to a single person. Of course, there is a sense in which we discount moral responsibility over time in some cases (such as in the criminal justice system), but it is commonly held that persons are temporally extended agents, and that any action an agent makes in his life can be properly attributed to a single agent.

As Parfit foresees, many of the implications that he argues follow from reductionism for morality are disturbing. This is because, on his view, once we shift from believing that an agent is a unified being, from birth to death, to seeing that the varying degrees of psychological relations within a life can constitute different agents, we will formulate a radically different picture of agency and, in so doing, complicate our traditional notions of blame, commitment, etc. Similarly, once we see that the separateness between persons does not matter, things such as distributive principles will correlatively matter less. Parfit says that the unpleasant changes in our moral views might be viewed by some as a reason for showing that reductionism is false. But, says Parfit, "The truth may be disturbing . . . If some truth is

disturbing, this is no reason not to believe it. It can only be a reason for acting in certain ways."[68] Put simply, Parfit is arguing here that the disturbing nature of his moral conclusions does not give us a reason not to believe the metaphysical truth of reductionism. And implicit in his argument, which we have traced throughout this work, is the conditional that allows the metaphysical truth to entail moral truth. So, Parfit reasons, if we do not have reasons to deny the metaphysical picture of reductionism, we do not have reasons to deny its moral implications. But we can, in fact, assess Parfit's reductionist moral conclusions on their own terms, and see if the moral consequences of reductionism are convincing enough to reject our traditional moral beliefs, as they are structured around persons.

In what follows, I will examine Parfit's case for a moral theory that shifts the focus from whole lives and the separateness of persons to the facts of the psychological connections that persons comprise. I will draw from Parfit's own discussions on specific areas within morality and the law. Parfit's discussions of commitments and desert/guilt bring to the forefront Parfit's (MP2), while his discussion on changing distributive principles brings to the forefront (MP1). I will examine both of these premises and how these premises can be sorted out in real moral situations. This will be done by explicating what the major debates are between Parfit's reductionism and

[68] Parfit (1984): 324.

defenders of common sense morality. Ultimately I will draw a parallel between this debate and the debate we examined in Chapter Three between Parfit and Hume in order to put into perspective and assess Parfit's position.

Rejecting the Unity of Agency

Commitments and the Language of Successive Selves

When someone makes a binding commitment, we hold this person accountable for fulfilling this commitment regardless of how he may change over time. Similarly, when we make commitments to other people, how we may change over time does not factor into our original commitment. This is because we generally believe people to be single agents over time. This gives us reason to feel bound by past commitments to other people.

If we accept reductionism, it becomes less plausible that one should stick to a commitment made in the past. If a man makes a marriage commitment at age 20 to love and protect his partner, does this same man still have to stick to this commitment at age 80, when he is no longer strongly connected to his former self? According to what Parfit calls the "Extreme Claim" about commitments, once we accept reductionism, we have no reason to honor past commitments.

Commitments involve personal identity twice: the identity of the commitment-maker and the commitment-receiver. If there is sufficient weakening of psychological connections between the commitment-maker at t1

and t2, we can argue that this changes the nature of the maker's obligations. But a commitment involves another person – the recipient of the commitment. One may argue that as long as the recipient of a promise asks for promises made in the form, "I shall help you, and all of your later selves," then these promises "cannot be held to be later undermined by any change in my character, or by any other weakening, over the rest of my life, in psychological connectedness."[69] So, it is plausible that while the maker's obligation can weaken over time with the weakening of psychological connections, the recipient of the maker's promise can argue that the promise must still be carried out, since it was made to him and his future selves.

But there may be cases, Parfit argues, where one can regard oneself as committed to the former self to whom he made a commitment, regardless of this person's later self's wishes. One example is Parfit's case, "The Nineteenth Century Russian." In this case, a young Russian socialist plans to impart to peasants the estates he will inherit several years later. But this young man knows that over time his ideals may change radically. To guard against revoking his commitment later in life, the man signs a legal document to give away the estates, which can only be revoked by his wife. The man then makes his wife promise not to revoke the document. He says to her, "I regard my ideals as essential to me. If I lose these ideals, I want you to think that I cease

[69] Parfit (1984): 327.

to exist. I want you to regard your husband then, not as me, the man who asks you for this promise, but only as his corrupted later self."[70] When he asks her to revoke the document in his conservative middle-age, the wife refuses, feeling committed to her husband's former self. In this example, it is plausible for the wife to maintain her past commitment to her husband's former self.

Ultimately, Parfit argues that one can justify giving up a commitment if the self who made that commitment is sufficiently weakly connected to him now. Similarly, one can be justified in maintaining a commitment to a past self of someone else even if this person changes dramatically. Both of these claims turn on the language of successive selves. All that matters, for Parfit, are what self committed a promise and to what self this commitment was posed.

Adams argues that Parfit's intuitions and verdict in the "Nineteenth Century Russian" case turn on factors that are extraneous to the fact of successive selves. Adams argues that what is turning our intuitions in favor of approving the wife's refusal to revoke the document is based upon the moral value of the contemplated actions and the wife's judgment of their value.[71] Adams has us imagine a variation on Parfit's example, involving a young Russian who is an archconservative and who has his wife promise never to revoke the legal document that specifies that none of his estates should ever

[70] Ibid.

be given away to peasants. Over time, however, he and his wife become more liberal, and he asks his wife to revoke the document to reflect these present ideals. Do our intuitions lead us in the same direction as they did in Parfit's case? Adams argues that in this case, it is less obvious why the wife ought to feel obligated to uphold her promise to her husband's young self, partially because the content of the commitment is one that changes our intuitions on the matter. Whether or not the wife is committed to uphold her commitment, Adams argues, turns on the character of the commitment and her attitude toward its character, not any commitment to her husband's former self.

Adams's argument does not get us far. Ultimately the debate between Adams and Parfit is a competition of dueling intuitions, and, as such, Parfit can deny Adams's claim by digging in his heels and arguing that our sympathy with the husband's earlier ideals does not change the wife's obligation to her husband's former self. And even if this was the case, it is no affront to Parfit that reductionism has moral consequences that are counterintuitive. If we have structured our beliefs about commitments around the notion of personal identity, then of course any change to these beliefs made to reflect reductionism will seem unpalatable.

The objections from intuition to Parfit's view take a back seat to some more primary objections. The power of a promise lies in the idea that a

---

[71] Adams (1997): 272.

promise is something that will be maintained in spite of the potential psychological changes that might be incurred by the promise-maker. The reason we ask someone to make a promise is that by doing so, the person is saying that the promise will be upheld no matter what – and the "no matter what" seems to guard against psychological change, among other things. Moreover, if we did believe that a promise was justified in being upheld according to the strength of Relation R between the self who made the commitment and that person's current self, it seems like we would not have to make promises at all. Promises, on this view, would merely be *confirmations* someone could make about himself as he is presently constituted.

So Parfit's conception of commitments as varying in strength according to the weakening of psychological connections does not seem to capture the very nature of commitments. It should be noted here that this objection differs from Adams's because it does not claim that reductionism is wrong to accept because it is counterintuitive. Rather, it is the objection that Parfit is changing the very definition of a commitment to fit his reductionist picture. If accepting reductionism would have such consequences for the nature of commitments, especially to the extent that we end up changing what the term means, Parfit might just want to give up entirely and accept that the notion of a commitment is one that only makes sense if we structure our moral concerns around persons rather than psychological relations.

The nature of commitments, then, seems inherently bound up with the construct of persons. The object of a commitment (for instance, later compensation, love, etc.) requires its committer to place this object in some kind of temporal context. The difference between, say, a person compensating someone immediately and a person making a commitment to compensate someone is that in the latter case, the commitment depends upon a kind of temporal extension. Now, what makes this temporal extension important to the nature of a commitment is that it is implied that temporal extension might, in fact, make it the case that a person changes significantly. This is one of the implications of the temporal extension of persons. And as we discussed above, if we were to isolate the notion of a commitment from the notion that this committer will change over time, we would have stripped the notion of commitment altogether.

There is a stronger claim that comes out of this temporal extension element of our notion of commitments. This claim is that when we understand persons as being extended over time and essentially remaining the same person over this period of time, we see that persons become treated as objects, insofar as they can be targeted by a commitment. When someone makes a commitment, they make a commitment to *someone*. Therefore, when making a commitment, there is a targeted object. On Parfit's view, the correct object to base our commitments on is a smaller temporal slice of a person, demarcated

from other temporal slices of the same person by the degree of psychological connectedness. Our commonsense view tells us that the correct object of a commitment is a person as we normally conceive him/her – a person who survives psychological change over time. I have argued that the very notion of a commitment is only intelligible if we understand a commitment being upheld despite psychological change on behalf of either the committer or the person-object of the commitment. And because the notion depends on the committer and person-object to survive temporal change, the committer and person-objects of our commitments have to be persons, as they are normally conceived.

This notion of commitments ties in with Hume's emphasis on the way in which the passions essentially depend upon picking out persons as their objects. Commitments are motivated by the passions, such as the feeling of indebtedness, love, guilt, etc., and the passions must have objects on which to focus in order to be realized. In the case of commitments, it is essential that the object a commitment picks out is a person.

Parfit argues that commitments could plausibly shift to take successive selves as one's objects, but commitments imply that something is being upheld to someone else in spite of whatever change the commitment-maker incurs. Of course, this does not mean that the commonsense view of commitments, as they are currently structured around persons, is morally

infallible. Commitments are regularly broken or weakened. But when one makes a genuine commitment, s/he at least intends to keep it, and recognizes that by making a commitment, s/he is expressing an intention to do so in spite of any changes in their desires. Ultimately, commitments depend on fixing a certain referent, and referents must be persons in order to maintain the very meaning of a commitment. If we scale back to make successive selves the objects of commitments, commitments amount to little more than mere affirmations, and to call these "commitments" is misleading.

Implicit in Parfit's argument in favor of adopting a view that changes commitments to reflect selves rather than persons is the idea that commitments are, to some extent, arbitrarily focused on persons. Parfit is assuming that there can be multiple candidates for the referents of a commitment, and that reductionism just gives us reason to opt for selves, rather than persons, to be these referents. But, as Hume argued in Book Two, we see that commitments are motivated by passions, and passions are necessarily person-focused. Commitments, then, are not conceptually malleable, at least in this respect. In order to uphold commitments, we must retain our concept of persons, or accept the Extreme View about commitments.

In what follows, I will examine real cases where the issue of whom we ought to commit ourselves to arises in legal practice and public debate.

Commitments and the Law: Honoring Prior Directives

Parfit's account of commitments and selves has drawn attention in the legal literature concerning whether or not to honor prior directives authored by now-incompetent persons. This is a domain in which some of Parfit's moral insights have received serious consideration. Consider the U.S. Supreme Court case, *Cruzan v. Harmon* (1989). In 1983, Nancy Cruzan was seriously injured in a car accident, leaving her in a permanently vegetative state. Cruzan did not execute a living will, but did mention to many friends and relatives that she would like to have her life terminated if she were ever unable to lead a normal life. Her parents asked her hospital to remove Cruzan's life-sustaining nutrition and fluid tubes because of their belief that their daughter would have preferred to have her life terminated at this point.

The issue in this case is whether Cruzan, before her accident, ought to be considered an agent whose opinion to have her life terminated, if ever her life ceased to be worth living, still has purchase over her now-incompetent self. Some have adopted Parfit's reductionist view and have argued that the weakness of the psychological connectedness between Cruzan-now and Cruzan-prior justifies the demarcation of two different agents. Because Cruzan-now and Cruzan-prior are so weakly connected, they can be properly considered two different people. As a result, there is no compelling reason

why Cruzan-prior has any relevant say when discussing the future of Cruzan-now.

Rebecca Dresser (1986) uses Parfit's successive selves account and his stance on the nature of commitments to argue against the use of prior directives, or living wills, to determine how a now-incompetent patient ought to be treated. Because Parfit argues that a sufficient weakening or complete loss of psychological connections between two selves might justify thinking of these selves as different agents, Dresser argues that there is no reason to treat the interests of a person's formerly competent self as more important than the interests of this person's current, now-incompetent self.

Dresser's argument echoes Parfit's treatment of his imaginary Russian socialist. If there is little to no psychological connectedness between the competent individual and this individual's incapacitated or altered state, Dresser concludes, "then there is no particular reason why the past person, as opposed to any other person, should determine the present person's fate."[72]

Nancy Rhoden (1990) rejects Dresser's Parfitian argument against living wills on two grounds. First, she argues that incorporating the language of successive selves into our moral theory threatens to wreak social havoc. Rhoden argues that reductionism threatens to swell our moral theory with a proliferation of new agents, which not only undermines our ability to target

[72] Dresser (1986): 380-1.

specific individuals but also makes this targeting less and less precise. This is because different people can claim not to identify with their former selves according to different and personalized criteria, and still the law's objective criteria for weighing psychological connectedness might confer a different decision. As Rhoden remarks, "the principle 'one body, one person' is a virtual necessity for the criminal justice system, for duties to honor one's contracts, or to pay for one's torts. Without unified personal identity, 'new persons' could spring fully formed into existence and legitimately could deny all family and financial obligations."[73]

Rhoden also argues that it would be especially peculiar to use Parfit's account of successive selves to make a case against the use of prior directives. The point of executing a living will is to identify oneself with a future person, which presupposes that there is a morally relevant connectedness between these two stages of one's life that transcends Parfit's Relation R. In cases where a person becomes incompetent and there is a competition between this person's current interests and former interests, Rhoden argues, it is not usually the case that this person has changed his mind, but rather that the person "has become too incapacitated to have a mind to change. A more appropriate question might be not whether the incompetent is the *same* person, but whether, in moral terms, he is a person at all."[74] According to Rhoden,

[73] Rhoden (1990): 854.
[74] Ibid.

considering a now-incompetent person as a kind of agent is misguided, and that in these cases, we ought to honor prior directives because these persons' formerly competent selves are, in a sense, taking care of themselves in a different state, not trying to exert power over an autonomous agent.

Rhoden and Dresser are arguing about the proper basis of our moral practices, and their debate is one that makes sense in the context of the debate we have examined between Hume and Parfit. Parfit argues that a correct view of agency is one that reflects the degrees of Relation R over time. On his view, two agents can properly be demarcated if there is a sufficient weakening of Relation R, and as a result, we must give up the idea that a former self has any authority over a future agent. The fact that these agents come into existence at different times and exist in essentially the same body is a trivial point. What makes this trivial is that personal identity has no metaphysical basis in and of itself, apart from its underlying psychological and physical facts.

So, we return to two of the implicit Parfitian premises we elucidated in Chapter Three, as well as Parfit's conclusion:

(P2) The only *genuine* facts are those that are metaphysically basic.

(P4) Personal identity does not *in itself* have metaphysical reality apart from that of the psychological and physical facts upon which it supervenes.

(C) Therefore, only the psychological and physical constitutive facts are metaphysically real.

It should be clear now that Dresser adopts the structure of Parfit's argument to reject the appeal to prior directives. Because all that matters are the psychological facts, and because the psychological facts obtaining between Cruzan-now and Cruzan-prior are so weak, we can properly think of these two women as two separate persons, and hence deny that Cruzan-prior has any authority over Cruzan-now.

Rhoden argues that agency cannot simply be sliced up according to varying degrees of psychological connectedness. When we reduce persons to these constitutive particulars, we come to believe that agency can be extracted from the whole of a single life and be justifiably be reduced to its psychological connectedness to future or former selves dictates. There is more at issue than just the degrees of psychological connectedness between two selves; rather, the crucial factor in these cases might be an authorial stance that individuals take to their lives. The sense that a person can identify with future selves and express his desires about how these selves ought to be treated if they become incapacitated might be a crucial relation that cannot be captured merely by comparing the psychological similarities between two selves. Parfit's view of agency fails to take account of authorship, or the way in which persons are able to identify, to some extent, with earlier stages of themselves, or project into the distant future. Rhoden is arguing ultimately

that practices such as developing prior directives are not only reasonable, but genuinely reflect authorship, an important feature of personhood. The Parfitian cannot make sense of authorship on a view that scales back agency by not allowing a single agent to incur substantial psychological change.

In many ways, Rhoden's argument can be interpreted to be an echo of Korsgaard's unity of agency argument in Chapter Two. But while Rhoden, like Korsgaard, similarly emphasizes authorship and how authorship makes sense of change over time, Rhoden's argument for authorship does not have to fall into the same question-begging trap that Korsgaard's argument falls into. Korsgaard argues that the unity of agency is a metaphysical necessity derived from Kant's unity of consciousness claim. As we saw in both Chapters Two and Three, claims to this extent ultimately end up susceptible to Parfit's reductionist aims, since they grant that the metaphysical facts of the matter are what is morally important, and that these ought to shape agency.

Rhoden's argument is more strongly bolstered by an appeal that Hume makes to the naturalness of our construct of persons without making this appeal tethered to a metaphysical argument. Even if Parfit is right in pointing out that psychological relations come in degrees and hold between different selves, he cannot argue convincingly that these degrees of Relation R have moral significance. Hume's appeal to the naturalness of the fiction of personal identity is an argument that precisely rejects any appeal to metaphysics in

morality. By rejecting this premise in Parfit's argument, Hume is not vulnerable to the question-begging trap Korsgaard landed herself in with an appeal to the necessity of a unity of agency.

I will now consider another example in the moral sphere where Parfit allows psychological connectedness between stages of a person to dictate agency. In the case of desert, Parfit claims that any agent not closely connected psychologically to the criminal who committed the crime ought not to be punished for the criminal's misdeeds. In this following section, I will frame the debate around the debate between Hume and Parfit and show that Hume's account of agency allows us to make proper sense of desert.

Desert

On the traditional view of desert, persons who commit crimes ought to be punished for their wrongdoings. On this view, persons are unified, single agents who should always be held responsible for their wrongdoings and receive appropriate punishment, regardless of when they committed their wrongdoings and to what extent they identify with their former criminal self.

For the reductionist, we can accept either the Extreme Claim or the Moderate Claim about desert. According to the Extreme Claim, in the absence of this fact, there is no reason to punish someone just because he is psychologically continuous with his former criminal self. If we adopt the

Moderate Claim, we could argue that punishment is justified only if the individual being punished now is strongly psychologically connected to the individual who committed the crime.

To understand the reductionist and non-reductionist approaches to desert, consider a variant of the fission case. Suppose I commit a crime before my brain is divided and placed into bodies A and B. Who ought to be punished? According to Parfit, the non-reductionist would believe that whoever possesses the deep further fact of personal identity deserves to be punished. The non-reductionist, then, has four options regarding who will be identical to my former self: (1) I am neither A nor B, (2) I am A, (3) I am B, and (4) I am both A and B. Suppose a non-reductionist decides that I am A, due to the fact that A possesses some sort of metaphysical appendage that makes A a better candidate for being me. So even though A is me, does B deserve to be punished? On this view, even though B is psychologically continuous with me, still B does not deserve to be punished for my crimes, because he is not identical to me. Of course, Parfit argues, the non-reductionist must admit that psychological continuity matters in some sense.[75] For instance, we may have reason to detain B (but not punish her) if I was a homicidal maniac before my division, because presumably B would be to some extent similarly psychologically disposed as my former self. But this

[75] Parfit (1984): 324.

consideration and others aside, on the non-reductionist view, B is still not responsible for my former self's crimes, since only the deep further fact of personal identity warrants punishment in this case.

Of course, Parfit is sketching the non-reductionist position unfairly. For Parfit, non-reductionism is not merely a denial of Parfit's brand of reductionism. As discussed in Chapter One, Parfit argues that non-reductionism necessarily involves a belief in the existence of some metaphysical appendage (a "deep further fact") to warrant personal identity. On this interpretation of non-reductionism, then, the only way to make either A or B a candidate for punishment of my former self's crimes is to make it such that one of them has preserved the deep further fact of personal identity. So, in this thought experiment, Parfit assigns the deep further fact to A, knowing that doing so is supposed to appear arbitrary. So Parfit's sketch of the non-reductionist's stance in this case makes the non-reductionist look like he is desperately holding onto some dubious notion of a metaphysical appendage in order to preserve the notion of desert.

While Parfit's sketch of the non-reductionist's stance in this case is unfair and hardly constitutes a decisive rejection of the non-reductionist view, I will put these worries aside, so as to see how palatable Parfit's reductionist moral views are on his own terms. I will then pit Parfit's reductionist claims against the Humean argument, which, while non-reductionist, escapes the trap

into which Parfit intentionally forced his own non-reductionist by unfairly sketching the non-reductionist position.

For Parfit, the reductionist can defensibly argue for the Extreme Claim and the Moderate Claim. On the Extreme Claim, neither A nor B deserve to be punished. Without the deep further fact about identity, there is no desert. Desert, therefore, is incompatible with reductionism. The reductionist can also make the Moderate Claim, which states that because all that matters is psychological connectedness and continuity, both A and B ought to be punished. On this view, degrees of Relation R ought to matter morally in the case of desert.

Suppose that someone convicted for a crime now is only weakly connected to his former criminal self. In this case, according to the Moderate Claim, the convict's punishment ought to reflect this change in degree of Relation R. In other, more extreme cases, someone can undergo such a radical change that he is not psychologically connected whatsoever to his former criminal self, and in this case, we can argue that the person does not deserve to be punished. Parfit argues that this appeal to psychological connectedness to warrant desert might be a reason why we have statutes of limitation, which mark the temporal boundaries beyond which we can no longer punish a criminal.[76]

[76] Parfit (1984): 326.

It is important to note here that Parfit's claim that reductionism is a reason we have statutes of limitation, which appears to be a way of making his reductionist proposal appear more intuitively plausible, should not be accepted. Statues of limitation are enforced to discourage an unreasonable delay in bringing civil lawsuits and criminal prosecutions by specifying a time after which one can no longer file suit. Many practical reasons motivate these laws. For example, one major reason might be that after a certain amount of time, the testimony of witnesses might become increasingly vague and insubstantial, thus making it harder for the courts to entertain a fair and accurate trial. Statues of limitations are not motivated by the idea that agency can be determined by psychological connections, to the extent that we might fairly judge a person to be a different agent than his former criminal self. Just because the courts will not entertain a case against an 80-year-old Nobel Peace Prize recipient who beat up a police offer in his youth, does not mean that in the eyes of the law, this man is not the same agent who committed the crime. Parfit's argument that the law reflects his radical reductionist agenda in this case, then, is misleading.

Parfit argues that the moderate reductionist view of desert treats the future self of a criminal as a "sane accomplice."[77] So of some past crime, a person's desert ought to correspond to the degree of his complicity with his

[77] Ibid.

criminal self. Rebecca Dresser spells out Parfit's appeal to the concept of accomplices by examining how accomplices are currently treated under the law. Under the law, Dresser says, accomplice liability subjects the accomplice to conviction for the same offense, and hence, the accomplice is subject to the same range of punishments posed to the criminal. But within this range, the accomplice's sentence may depend on his individual blameworthiness, rather than the blameworthiness of the criminal.[78] In this way, reductionism is able to avoid the Extreme Claim by asserting that even a later self weakly connected to a past criminal is still implicated by his previous self's actions, but his present culpability for this self's crimes will also depend on the present self's current psychological attributes, such as whether or not he currently continues to entertain criminal inclinations.

Dresser argues that reductionism can be reconciled with desert. As she notes, it is already intuitively and legally sound that in order to justify punishment, "it is necessary to determine whether there is a morally relevant psychological connection between an offender and a person subject to punishment."[79] One type of justification for punishment is retribution.[80] The principle of retributive desert asserts that punishment is justified because once a criminal has committed a wrong against society, society is justified in

[78] Dresser (1990): 425.
[79] Dresser (1990): 427.

proportionately harming that person. Retribution is not concerned with yielding a future good in punishing a criminal, and only seeks to punish the very person who committed the crime. Reductionism threatens to undermine retributive desert because retributive desert relies on the concept of a person enduring over time. In order to punish a criminal, the law must treat the person who committed the crime as a single unit over time who will serve to be a target of societal blame and who will presumably persist over time to regret his former actions.

Dresser argues that reductionism pinpoints what lies at the heart of this general idea. Relation R, Dresser contends, encompasses reasons for holding someone guilty for a crime, and hence does not undermine retributive desert. Since the law also holds that psychological connections between the criminal and the current convict are morally relevant, all reductionism does is make it incumbent upon us to refine our moral beliefs to reflect what connections are relevant and the varying strength of these connections over time. For the reductionist, punishment is defensible only if the person punished is psychologically related enough to the criminal. This judgment, Dresser admits, is difficult to enforce. To do so, we must go beyond Parfit's work to explore which connections are important ones and which psychological features must necessarily obtain in the criminal's later self to explain why he

---

[80] Dresser (1990) notes that there exist four classic justifications: retribution, deterrence, incapacitation, and rehabilitation (420). I will only be concerned here with retribution, as its

might be excused from punishment. Currently, any judgment about punishment would rest merely on speculation and could be indiscriminately imposed.

Dresser argues ultimately that while reductionism seems to place difficult demands on the criminal justice system, it still reflects what is ordinarily the case in our moral intuitions and what is already echoed in the law. In most criminal cases, a sufficient number of psychological connections are present to compose psychological continuity and hence justify desert. Similarly, even many direct psychological connections between the convict and the criminal exist. Also, there are cases that exist where sufficiently weak connections over time justify changing the convict's degree of punishment.

Reductionism thus serves as an implicit foundation of our justification for punishment. Reductionism's further task, then, is to bring clarity to decisions that must be made in cases where the degrees of psychological connections between convicts and criminals are not obvious. Indeed, Dresser argues that reductionism "counsels a more open and systematic examination of the possibility" that a later person might deserve less punishment.[81]

A crucial weakness in Dresser's argument is her failure to distinguish between what reductionism can practically provide us and what it

---

defense is analogous to defenses of these other justifications.
[81] Dresser (1990): 435.

theoretically provides. A reductionist asserts that personal identity is not always a determinate, all-or-nothing matter, and, as such, there can exist degrees of connectedness and continuity between two stages of a person. Dresser argues that this feature of reductionism is one that can practically alter and help precisify our understanding of criminal justice. But if reductionism is to do so, these varying degrees of connectedness and continuity must be not only accessible but also quantifiable. Otherwise, they cannot inform decisions regarding punishment of the kind that Dresser says they can inform. Theoretically, it seems possible that connectedness and continuity might come in degrees, but it is another case entirely to determine how these varying degrees are in fact realized.

One way of evaluating Parfit's reductionist proposals is to examine what notion of personal identity we use in moral judgments and in the legal system. Recall Baier's focus on the perception of human bodies, and how this plays a crucial role in structuring our moral beliefs. Baier's emphasis on bodies is to show how personal identity is a notion we utilize in making practical judgments. We can wax philosophical "in our armchairs"[82] about what personal identity is and how most of our notions of personal identity hinge on psychological connectedness and continuity over time, but this will not necessarily yield for us any practical way of altering our moral judgments.

[82] Baier (1991): 138.

As Baier argues, the perception of persons as human bodies play a crucial role in practical morality. We see others as single-bodied persons and see ourselves through the eyes of others. Baier argues, “Since our conception of a fellow is of a flesh-and-blood person, then whatever conception we have of the identity over time of a living expressive body will be the core of our notion of a person’s identity, and so of our own identities.”[83]

Being embodied, then, reinforces the idea that persons are real entities, and the idea that having a single body that changes but remains essentially the same matters. Parfit argues that having a body is not a deep fact. Parfit is able to generate a number of counterfactual scenarios that show that in these special circumstances, we need neither physical continuity over time nor the uniqueness clause of personal identity to preserve what supposedly matters. But whether or not the importance of these two notions falls apart in counterfactual scenarios does not detract from the plain fact that this notion of personal identity has structured and is reflected in our moral practices. At the heart of our moral practices and judgments lie the passions. The notions of desert, justice, retribution, and so on all stem from our most primal passions. And as Hume argued, the passions actuate persons.

Integral to wielding the concept of persons is an understanding of how, first, we are different bodies than other bodies, and how, second,

---

[83] Ibid.

understanding this differentiation gives rise to the passions and certain other attitudes indebted to interpersonal dynamics. For example, the notion of envy would be unintelligible if we were not able to perceive other bodies and see these as persons. Without persons, we would be bereft of an object for our passion, and envy, then, would not make sense. As Baier argues, "Since we can see the separateness of human bodies . . . we know what makes one person different from another, and the experience of disagreement and conflict will reinforce that knowledge. Equally we know about the inter-dependence of flesh-and-blood persons for normal growth, for belief formation, for self-knowledge and for the sustaining of pride."[84] So, on this view, simply perceiving other bodies as persons, and perceiving oneself as a person reflected by these other persons, is the foundation of our basic moral practices.

Now, return to Dresser's defense of reductionism in the case of desert. Dresser and Parfit both argue that reductionism captures what matters about personal identity, and as a result, should be used to reshape our moral practices. They argue that the reasonableness of a reductionist account of personal identity arises when we consider counterfactual scenarios and are able to strip away notions of unique embodiment and physical continuity from what seems to be the most crucial elements of personal identity. But as I

[84] Baier (1991): 136.

argued above, whether we can conceptually peel these notions away does not make it such that a revisionary approach toward morality is in order. This is because we can have a practical conception of personal identity that is justified and defensible, independent of a purely theoretical conception of personal identity.

A theoretical peeling, taken on its own, may discard precisely what is most relevant to our actual moral practices and intuitions. As Hume argues, we can see persons in two very different lights.[85] Upon skeptical examination, we can see our notion of persons as bodies as being groundless. But pragmatically, we can see how natural these notions are, and how they are essential to our very survival. Does the fact that we have two very different notions of personal identity make it the case that we have to dispense of one? Parfit believes so. But we have yet to see why our pragmatic view of persons is somehow flawed, in spite of its utter naturalness and the foundational role it has played in moral thinking.

Parfit uses an appeal to intuitions to ground the reasonableness of restructuring our moral practices to reflect reductionism. After all, Parfit has been painting reductionism as the ultimate defender of "what matters." The conclusions Parfit draws from his counterfactual scenarios are supposed to illuminate what lies at the heart of our moral thinking. Dresser uses this

[85] Hume (1978): T 253.

appeal to her advantage as well. She argues that our legal notions of desert already allow for varying degrees of psychological connectedness to dictate different degrees of punishment. But as I argued against Parfit's appeal to the legal philosophy behind statutes of limitation to justify his reductionist moral views, her appeal to current practices as somehow validating reductionism is entirely misleading. Punishment varies, not in terms of psychological connectedness *per se*, but rather to reflect a person's ability to identify with his former criminal self, recognize that his action was wrong, and give the courts reason to believe that he will no longer pose a threat to society. Changes in psychological connectedness between the criminal and the current person may be a redundant justification of how punishment may change as a criminal learns from the error of his ways.[86]

But even if the changes in psychological connectedness in the strict reductionist sense do seem to be an additional justification of how legal punishment works, that does not mean that the radical reductionist claims that Parfit wants to advance are appropriate. Indeed, simply noting that psychological change over time does factor into degrees of punishment does not mean that the law considers the former criminal and this criminal's current self to be different agents. In fact, the opposite seems to be true: Punishment is made less severe in cases where the criminal's current self is able to

[86] For further examples of redundant justification, see: Johnston (1992): 595-596.

identify with his former self, and genuinely believe that he has changed for the better. In this sense, the criminal's current self understands that he is the same agent who committed crimes, but that he has been able to change himself substantially and develop new attitudes and ideals to supplant the old ones. Our moral practices and the law seems to reflect this notion because there is a sense in which we feel it is important for a person to have changed himself. It is not as important that we have simply discarded a disagreeable agent for a new, law-abiding agent. Rather, it is important that an agent can change and later repent, and this can only be done if the agent can identify with his former self and his actions.

In this section, I have evaluated Parfit's argument that we ought to abandon our commitments to the idea that persons possess a unity of consciousness, which makes them single agents over time. This comes to the forefront when Parfit examines how reductionism ought to change our views about commitments and desert. The notion of a commitment and the notion of desert are structured to reflect the unity of consciousness by treating persons as single entities over time. Parfit says that an illumination of what matters to us in morality shows that this view is problematic and does not matter. Instead, Relation R matters, and this is what ought to structure our moral practices.

I have brought the reductionist views of both commitments and desert into perspective by pitting the reductionist's essential claim against the Humean position. The Humean position argues directly against Parfit's conclusion, that only the constitutive facts of personal identity are metaphysically real. Hume argues that our concept of personal identity is something that plays an important role in our moral practices, and in so doing, is not merely reducible to its constitutive facts. We do not rely on metaphysical elucidations of our moral practices to justify them prior to acting. In this sense, our moral practices have practical utility and are not touched by Parfit's arguments. I have also argued that Parfit's appeals to our current practices as somehow reflecting the heart of the reductionist picture are misleading. Instead, many of Parfit's reductionist views may only be redundant justifiers of certain practices and laws, not their critical motivators. It is important to keep an eye to Parfit's attempts to make reductionism more palatable by appealing to current practices, because by accepting his claims, we are overlooking some of the more radical ideas imbedded in his reductionist agenda.

In what follows, I will examine Parfit's claim that we ought to reject the notion that the separateness of persons is morally significant. Parfit argues that the separateness of persons is not metaphysically deep, and thus should not bear on our moral thinking. He brings out this claim by specifically

focusing on distributive justice. In a way analogous to my arguments in this section, I will get at Parfit's essential reductionist claim, and pit it against Hume's position in order to determine which view is more plausible.

Rejecting the Separateness of Persons

Distributive Justice

The fact that persons are separately existing entities, each with his/her own life to lead, is less deep on the reductionist view. If all that matters is the persistence of psychological relations over a lifetime, and not identity, the fact that a person is a metaphysically distinct entity from another should not factor into our moral considerations. In this way, reductionism supports undermining the separateness of persons in the moral sphere. While the claim that the separateness of persons does not matter touches upon many potential moral questions, it most visibly comes to bear in the moral debate about the principles of distributive justice.

The reductionist focus on undermining the separateness of persons for the purposes of distributive principles is pitted against the Principle of Equality, which maintains that happiness ought to be fairly shared between different people, and that it is bad if some people arbitrarily receive fewer benefits than others. Imagine that in order to bestow a large benefit on one person, all I have to do is impose a smaller burden on another. On one view,

where the separateness of persons does not matter, one might believe that all that does matter is increasing the net sum of benefits, impartially considered.

This is one way of interpreting the reductionist view. On this view, I am morally obligated to impose the burden on the one person in order to increase the net sum of benefits. Of course, defenders of the Principle of Equality maintain that it is not fair to impose burdens on people for the sake of happiness for another. This might be motivated by a number of reasons, one of which would be that there is no necessary compensation. So for those who defend the Principle of Equality, the separateness of persons is a deep truth, and as a result, it is important that each person have at least an equal chance of receiving the same amount of benefits as another.

Parfit argues that because most of us are non-reductionists, the Principle of Equality appears to trump the reductionist view. This follows, of course, because as non-reductionists, we believe that there is a deep further fact involved in personal identity, and that thus the distinction between persons is important. What happens if we cease to be non-reductionists? Parfit argues in favor of widening the scope of distributive principles, but giving these principles less weight. In so doing, reductionism would succeed in undermining the separateness of persons.

First, let us set the stage for considering what it means to alter the scope of distributive principles. This is important because reductionists have

introduced and made crucial the notion of selves to their revisionary proposals for morality. A reductionist holds that the degrees of psychological connectedness within a life, yielding selves, are much like the divisions between persons. Incorporating selves as units deserving of moral consideration is central to the reductionist's first argument: persons do not possess a unity of consciousness that secures identity over time, and, as such, personal identity does not matter.

In the previous section, I granted the reductionist claim that selves can be relevant units in moral discourse, and I examined whether selves could be the proper recipients of moral consideration in cases of desert and commitments, without improperly tampering with the essential meaning of these terms. Now that we are turning to the reductionist's second argument – that the separateness of persons does not matter – it remains a question of whether selves are the proper recipients of distributive principles. It seems that if the reductionist is committed to a metaphysical claim about the importance of selves in a moral discourse, distributive principles ought to reflect these units.

But, this would make Parfit's rejection of the separateness of persons, while granting the separateness of selves, entirely arbitrary. Why ought we undermine what appears to be a natural separation between persons in favor of the less clear separateness of selves? Parfit sees the problem here

for the reductionist, and argues that we must choose one of the two following alternatives: (1) We must make it such that distributive principles apply both to separate persons and separate selves (thus widening the scope), or, (2) We must reject distributive principles entirely.

Parfit argues that the reductionist can plausibly widen the scope of distributive principles, thus granting the first alternative, but then deny that distributive principles have any weight. For Parfit, once we accept the notion of selves within a life, one must bite the bullet and admit that if distributive principles can be applied to different persons, they must also apply to selves. But widening the scope of distributive principles in this way does not entail that distributive principles have moral weight.

To see how this would work, consider the following example. Suppose we are deciding whether to impose a burden on a young child. Doing so will either result in benefiting the child later in life, or it will result in benefiting another child. Does it matter morally which alternative results from imposing the burden? By giving distributive principles more scope, the reductionist can argue that not only would it be wrong to harm this child for the benefit of another child, but also harming him for the sake of his own future benefit is equally wrong. In this case, the reductionist considers the child's future self like a different person, and therefore, both of the potential results from imposing the burden on the child are morally the same. Both of

these options fall under the Claim about Compensation, which states that a burden cannot be compensated by benefits to someone else.[87]

Now, while the reductionist cannot deny this claim, he can deny that this claim has weight. Compensation presupposes personal identity, and since the reductionist believes that personal identity is less deep, the Claim about Compensation is less morally important. To illustrate how the reductionist can both grant widening the scope of distributive principles and deny their weight, Parfit returns to the nation analogy. Parfit argues that a reductionist about nations never denies that nations exist. They just deny that the existence of a nation involves anything over and above a nation's citizens and their actions. So while the reductionist sees that there is such a thing as a nation, constituted by lower-level facts, he can also maintain that the existence of a nation does not in itself matter. Analogously, in granting distributive principles more scope but less weight, the reductionist is granting the fact that separateness between persons and selves do *exist*, but denies that the separateness between persons and selves is *important*. Because these separations are not important, reductionists can still claim that they can describe lives impersonally by just focusing on the psychological and physical facts of a person rather than focusing on them as unified beings over time.

[87] Parfit (1984): 337.

Parfit's analogy between his view about distributive justice and nations brings us back to the tension we observed in Parfit's reductionist analysis of nations in Chapter Three. As we saw earlier, Parfit holds that personal identity, like nationhood, is just reducible to its constitutive base. But Parfit makes this claim about the utter reducibility and unimportance of identity by simultaneously granting a kind of reality to personal identity. The tension in Parfit's account of personal identity lies in the fact that the constitutive facts that make personal identity real are the same facts by which personal identity can be deflated. It is unclear, then, what status we are assigning personal identity when Parfit grants that it both exists in virtue of, but yet is *wholly reducible* to, its constitutive base. Parfit seems to be intentionally vague about what existence is when he says that reductionists "do not deny that people *exist*"[88] but could, in fact, "give a complete description of our lives that was impersonal: *that did not claim that persons exist*."[89]

Parfit seems to argue at some places that persons exist only because of the way we talk. So Parfit is saying, then, that persons exist as the notional referents of words that are bandied about in conversation, but are not metaphysically real. But even if this is the case – that persons exist only because of the way we talk – it is not sufficient to say that this fact alone

[88] Parfit (1984): 341.
[89] Ibid. Emphases mine.

makes persons unimportant constructs. Parfit seems to be relying on the rather implausible picture that we have arbitrarily incorporated person-talk into our language, and that this person-talk is some kind of strange linguistic accident.

Parfit fails to explain, first, why we talk about persons, and second, why person-structured discourse seems to be important. Hume, on the other hand, can explain these two points by appealing to his fictionalist account of personal identity. Person-talk is not merely a linguistic phenomenon. Rather, on Hume's view, we create the fiction of personal identity as a result of the relations we perceive and conventions we employ. And fictions are, hence, entirely natural and reasonable constructs, and in turn make sense of the passions and, hence, moral life.

Conclusion

Parfit claims that accepting reductionism ought to change our moral thinking. This is because Parfit believes that reductionism properly elucidates the morally important constitutive facts of personal identity. We have unpacked Parfit's position to reveal his premise that the only genuine facts that ought to have moral purchase are those facts that are metaphysically real. The facts that determine our judgments about personal identity – including psychological connectedness – are metaphysically real, while personal identity, in itself, is not. Therefore, we ought to focus on the metaphysical

facts, and not personal identity. So, reductionism, having illuminated these facts, pushes for a moral discourse that shifts its focus from persons to Relation R. And in doing so, we see that there are two major reductionist proposals:

(1) The separateness of persons is not morally important, and

(2) Persons should not be regarded as single, unified agents over time.

In this chapter, I have focused on each of these proposals and examined three different moral arenas for which these claims have important implications. Parfit's first claim challenges our traditional views about distributive justice. Claim (1) denies that the separateness of persons is metaphysically deep. Even though there are different views about distributive justice, all of these views depend upon an essential and non-trivial separateness of persons. Claim (1), then, is a pressing claim which these theories need to address. Parfit's second claim challenges our views about commitments and desert, by claiming that these ought to reflect selves, demarcated by degrees of Relation R, rather than persons, as they are normally construed.

I have evaluated the reductionist position in the moral arena by pitting Parfit's reductionist proposals against the Humean notion of personal identity. In evaluating Parfit's first and second claims, I have argued ultimately that

our construct of a person is an independent fiction irreducible to its constitutive base. In being irreducible, we come to see how the construct of persons is reasonable and necessary for moral language. I have also illuminated the way in which Parfit draws himself into a trap by granting existence to persons but denying that the existence of persons matters. Parfit thinks that he can get out of this trap by gesturing toward the idea that persons exist only insofar as they are things we happen to talk about. But I have argued that the fact that we engage in discourse about persons is not simply a superficial fact about our language, but rather can be explained by Hume's notion of the fiction of personal identity.

In this sense, then, Hume's account of personal identity trumps reductionism by showing how the construct of personal identity can be independently reasonable and useful. Parfit's reductionist account also falters on its own terms because of the tension Parfit runs into by granting the construct of a person existence in virtue of its constitutive base, but then deflates its importance by its constitutive base. The inconsistency of Parfit's reductionist proposals, as well as the reasonableness of Hume's position, both show that Parfit's revisionary reductionist proposals for morality are far from decisive, and that the very construct of personal identity that Parfit argues against might be independently valuable for our moral discourse.

At this point, it is important to note that Parfit has in mind the following implicit conditional: If our metaphysical picture of persons is correct, then our moral views, insofar as they reflect the metaphysical truth, are correct. Parfit argues that his metaphysical picture is correct, and this granted premise, in conjunction with his implicit conditional, allows him to assert that his moral conclusions are correct. As I announced in the beginning of this chapter, my aim was to grant the antecedent of Parfit's conditional (that is, his metaphysical picture) and deny Parfit's claim that the metaphysical picture entails a proper moral view. I did precisely this by examining other moral views and arguing that Parfit's implicit conditional is far from decisive.

In addition to denying the entailment, I have challenged Parfit's ethical picture *on its own terms*. Even if we granted Parfit's conditional, Parfit cannot assert that his moral conclusions are correct unless Parfit's metaphysical picture of persons seems more plausible than denying our traditional, natural moral views. As I have argued, Parfit cannot reach this point, because denying our moral views is a much more implausible task than granting Parfit's speculative metaphysics.

In this Chapter, I have shown that our traditional moral views are justified and defensible. Parfit, then, has not made his metaphysical picture of persons plausible, or at least more plausible than the denial of our traditional

moral views, which, as we have seen, are clearly more defensible than Parfit anticipates when he undertakes his revisionary moral claims.

Chapter Five: Fictionalism and Minimalism

The last two chapters have been dedicated to explicating and defending what I believe to be the strongest case against Parfit's reductionist account of persons. This case rests on Hume's fictionalist account of personal identity.

Many objections launched against Parfit attack Parfit's claim that a true metaphysical picture of persons ought to have normative significance. Parfit makes two claims in this regard:

(1) Because there is no unproblematic metaphysical component that secures personal identity, personal identity does not in itself matter.

(2) The metaphysical components that constitute personal identity do exist, and because of their metaphysical reality, we ought to restructure our moral and prudential concerns and beliefs so as to reflect the metaphysics.

As we have seen, criticisms of both these claims are interrelated. These objections involve defending accounts of why personal identity is justified, insofar as it is a practical construct. But writers such as Korsgaard do not show how it is that we can preserve the construct of personal identity without grounding it implicitly in a metaphysical picture of personal identity by relying on an appeal to necessity. Hume's account provide a more defensible view of persons. On a Humean view, even if Parfit gets every

metaphysical piece of the construct of persons correct, this does not at all diminish the independent value of the construct of persons.

I will now defend Hume again by connecting Hume's fictionalist account to the debate between Parfit and Johnston, whose minimalist account of persons mirrors aspects of Hume's fictionalism. I will examine Johnston's account of minimalism as well as Parfit's response to Johnston. Ultimately, the weakness in Johnston's minimalist account is the claim that person-directed beliefs ought to be understood as being utterly basic and defensibly unjustified. Hume, on the other hand, shows that while these beliefs are utterly basic, they can be justified. By bringing Hume's account into accord with Johnston's minimalism, I reveal what I believe motivated Johnston's view, and which I take to be the central point of disagreement between Hume and Parfit. I conclude this final chapter, then, by bringing to the forefront this disagreement and arguing that Hume provides us the most justifiable account of persons.

The Supervenience Structure of Reductionism

To understand the power of the minimalist argument, it is important first to review briefly the structure of Parfit's argument. As we saw in Chapter 1, Parfit argues that the metaphysical facts of personal identity necessitate a change in our moral and prudential thoughts and practices.

Parfit claims not to deny that persons exist. He argues that the facts of personal identity just consist in the holding of certain mental and physical facts. So, the facts of personal identity supervene upon physical and psychological continuities. There cannot be a change made, then, at the level of the supervening fact of personal identity that does not affect the constitutive metaphysical facts. Of course, Parfit allows that it could have been the case that Cartesian egos or further metaphysical facts existed to compose the supervenience base for personal identity.[90] But in the absence of such superlative metaphysical facts, the only facts about persons that matter are those facts that constitute persons – and these facts are psychological and physical facts. Parfit's case for reductionism thus hinges on the idea that the supervenience base of personal identity is what matters above and beyond the mere supervening fact of personal identity.

Parfit follows this argument with the claim that our concerns, which are currently based on our non-reductionist notions of personal identity, should change in order to reflect the metaphysical facts. This can be summarized as the Grounding Assumption: The justification of our normative practices requires that they be grounded in facts about personal identity.[91] Given this assumption, Johnston argues, "the superlative entities can seem to be the only things that would confer the required privilege on our practical

[90] Parfit (1984): 227.
[91] I borrow this term from Perrett (2003): 375

concerns."[92] The presence of these superlative facts would have made personal identity more metaphysically "deep." So, had there been superlative facts, our normative practices might be wholly justified. Instead, these practices are simply the results of incorrect assumptions about our metaphysical composition. This is how we arrive at Parfit's Extreme Claim. Without superlative entities, we cannot justify the reasonableness of our practices.

Minimalism is, in effect, a rejection of the idea that our practices, as they are structured around personal identity, are unjustifiable in virtue of not being grounded in anything that is metaphysically real. We can see here that Johnston and Hume are in accord.

Minimalism

Minimalism is an account of justification. It rejects Parfit's Argument from Below, which we saw in Chapter 1.

The Argument from Below:

(1) If reductionism is true, personal identity just consists in certain other facts.

(2) If a fact consists in certain others, it is only these other facts that have rational or moral importance. We should not ask whether, in themselves, these other facts mater.

[92] Johnston (1992): 591.

(3) Personal identity cannot be rationally or morally important. What matters can only be one or more of the other facts in which personal identity consists.[93]

Johnston rejects the Argument from Below and replies with the Argument from Above. This can be generally sketched as follows:

The Argument from Above:

(1) If reductionism is true, personal identity just consists in certain other facts.

(2) It is *not the case* that if a fact consists in certain others, it is only these other facts that have rational or moral importance.

(3) Personal identity is morally and rationally important.

(4) The smaller facts, which personal identity consists in, are *derivatively* of moral or rational importance.

So, the Argument from Above rejects Parfit's (2) and (3) from the Argument from Below. By rejecting Parfit's Argument from Below, Johnston is rejecting the emphasis Parfit places on the mere absence of a Cartesian ego or another superlative further fact.

The minimalist position's substantive charges against Parfit's reductionism are clearest in Johnston's minimalist defense of self-concern. As we saw in Chapter Two, the common sense egoist holds that our special concern for our futures is justified in virtue of the fact that these future stages will be of one and the same self. Johnston defends this view by appealing to the naturalness of self-concern, and calls such concern "non-derivative." A

[93] Parfit (2003): 305.

review of non-derivative self-concern will provide us a good point at which to return to Hume's fictionalist account.

The Minimalist Defense of Non-Derivative Concern

For the minimalist, self-concern does not require for its justification the existence of superlative further facts. Rather, self-concern is something that is valuable within the greater pattern of self-referential concern. Johnston uses the term "self-referential concern" loosely. Self-referential concern includes our concerns about our futures, our families, our loved ones, our friends, and so on. A person exists as a part of a vast network of relationships, and identifies with the positions he assumes in these relationships (i.e., he can be a father, a colleague, an employee, etc.). In each of these relationships, he can possess non-derivative concern for the good of these relationships and the other people to whom he is bound by these relationships.

For example, he might non-derivatively care about his partner's recent onset of depression. He cares non-derivatively for his partner simply because his partner is suffering. His care is merely on his partner's behalf. He cares, then, not because he is necessarily concerned with how his partner's depression will come back to affect him negatively, or how it might negatively impinge on the structure of the family. Of course, these might be

additional reasons to care about his partner's depression, but the presence or absence of these additional reasons does not impinge upon the intrinsic nature of his non-derivative concern for his partner.

Each of us has non-derivative concerns for people in virtue of their relationship to us. What justifies non-derivative concern? Johnston says, quite simply, that non-derivative concern does not need justification.[94] To make this claim more defensible, Johnston appeals to certain types of concern we have that we do not feel pressured to justify. One might argue, for example, that one is justified in being concerned about something if this concern makes the world better. But what justifies the concern that the world should go better? Johnston is implying here that justifying our concerns always comes to a stopping point, beyond which it seems ludicrous to attempt more and more justifications to support certain basic, fundamental concerns.

Johnston argues that non-derivative concerns that are justified "are those which will continue to stand the test of informed criticism."[95] This means that there are certain beliefs that are all-encompassing and the naturalness of which makes it difficult for critics to ground defensibly a decisive base of criticism that does not already assume some of the premises that base the very belief system these critics wish to attack.

---

[94] Johnston (2003): 270.
[95] Ibid.

Johnston compares the need to justify non-derivative concern to skepticism about the external world. Belief in the external world is a fundamental belief that cannot be undermined precisely because this basic belief is so encompassing and plays a decisive role in our reasoning. In the same way that we don't have to justify our belief in the external world, we should not have to justify non-derivative self-concern.

This argument echoes Hume's distinction between the questions, "Does P exist?" and "What causes us to believe that P?" in his discussion of our beliefs about the external world.[96] But the difference between Hume and Johnston is that Johnston takes the second question, "What causes us to believe that P?" to be just as unanswerable as the first.

First, let us defend Johnston's and Hume's rejection of posing the first question. It is not that Johnston and Hume are just trying to fend off difficult questions about the justificatory nature of our most cherished beliefs. Their point is that we could not even pose the question without presuming certain notions that take for granted this belief in the external world. We take for granted that there is an existence outside of us when we ask questions. We take for granted the existence of an external world by implying that we understand the concept of existence, and that we must decide whether this predicate is appropriately applicable to "the external world." We could not have these concepts if we did not at least grant the existence of the external

world in some sense. It is for this reason that we should abandon the absurd question, "Does the external world exist?" and instead revert to the question regarding how it is that we think the external world exists.

This same claim about fundamental beliefs of the external world is also found in Wittgenstein's *On Certainty*, in which he argues, "I did not get my picture of the world by satisfying myself of its correctness; nor do I have it because I am satisfied of its correctness. No: it is the inherited background against which I distinguish between true and false."[97] Wittgenstein's idea that certain basic beliefs compose the "inherited background" against which we can make any statements is the picture of persons that Johnston wants to argue is true of our basic beliefs about self-concern.

Parfit clearly disagrees that our beliefs about self-concern can stand without justification. Indeed, a correct metaphysical picture of persons, Parfit argues, ought to justify any beliefs we have about persons, no matter how basic. According to Parfit, the absence of superlative further facts should displace self-concern. On this view, self-concern would be justified if and only if some realist view of persons was true. But, as Johnston rightly argues, our concern for others and for our future are never (or at least rarely) based on substantive metaphysical views. We do not justify our friendships, for example, on any deep beliefs about the nature of friendship or of our friends.

---

[96] See page 65 for this discussion.
[97] Wittgenstein (1969): prop. 94, p. 15e.

This is because we value friendship and other relationships non-derivatively, and, similarly, value self-concern non-derivatively.

Johnston is indicating a fundamental disjunction between our metaphysics and our values. The latter does not need the former for its justification. If anything, our metaphysical views might derive from our antecedent values. Our belief in Cartesian egos might be, at most, a redundant justification of self-concern.[98] For instance, if a person were to find out that Cartesian egos did exist, this person might then cite the existence of a Cartesian ego as a justification for having self-concern. But presumably, before he knew about the possibility of Cartesian egos, he already had self-concern, and this had non-derivative value for the reasons specified above. Since beliefs about one's metaphysical composition do not in general determine our values, how could Parfit convince us that a substantial change in our metaphysical views should have such normative force?

The demonstration that there is no metaphysical justification of our practices does not in itself constitute an argument against the legitimacy of these practices. As Johnston argues, Parfit simply fails to establish the appropriate link between pointing out what the metaphysical facts of personal identity are and making a case for how these facts would alter the concept of personal identity that structures our ordinary concerns.

---

[98] Johnston, (1992): 595.

Fictionalism on the Justificatory Status of Person-directed Beliefs

So far, Johnston's and Hume's accounts seem congruent. But Johnston's argument is missing a crucial element that we can profitably draw from Hume's fictionalist account. Johnston makes a case for the non-derivative nature of self-concern, but this, in a sense, amounts to little more than simple repeated declarations that some concerns do not need to be justified. This is not an argument. Johnston's account is vulnerable to Parfit, because Parfit needs only to deny the notion that we have some concerns that simply do not need justification. Repeatedly naming certain concerns that we do have and which are not justified does not get us anywhere with Parfit, precisely because Parfit already rejects the values to which Johnston appeals, as they are structured around personal identity.

Hume's fictionalist stance accomplishes what Johnston's minimalism does not. Fictionalism provides an account of how our self-concern is, in fact, justified, and shows how this justification need not be based on Parfit's metaphysical view. For Hume, persons are indeed facts that supervene on more particular facts about physical and psychological composition. But the facts about persons are not thereby reducible to their physical and

psychological constitutive facts. This is where Hume and Johnston agree. But Hume makes a decisive move that Johnston does not: the fiction of personal identity is justified because it is based on our understanding of certain relations, and conventions, which form to reflect to these. Relations and conventions in turn constitute the fiction of personal identity. And the fact that the fiction of personal identity is, then, essential to understanding the passions, makes the fiction of personal identity a reasonable one.

Johnston argues that we cannot justify self-concern. Hume agrees with Johnston insofar as Hume holds that our beliefs about persons (which are spelled out in our self-referential concerns) are beliefs that we take for granted. But for Hume, these beliefs are *induced* by certain things (as he makes clear in his distinction between the two questions we can pose toward our beliefs about the external world). Our beliefs about persons are first based on seeing certain relations, like causation and resemblance. We see the succession of different perceptions in an "inconceivable rapidity," such that these different perceptions become united via certain relations. More specifically, we see the succession of person-stages and can make sense of this succession by perceiving the relation of resemblance.

These relations then give rise to certain conventions about persons. These are practices that utilize the relations we perceive. As discussed in Chapter Three, we form the convention that future stages of a person are identical in kind to a person's present stage. Conventions about persons can

be more generally described as the ways in which we respond to the relations we perceive. It is a part of our response to the relation of resemblance over time of different person stages that we come to regard future continuers as being identical in kind to a person's present stage.

From these conventions, we arrive at the fiction of personal identity. Note that the fiction of personal identity, then, is one that is indeed grounded by certain relations. We can cite relations and conventions to justify how we arrive at the fiction of persons. This account of persons is stronger than Johnston's, simply because it can show what causes our beliefs about persons. It does not leave our beliefs about persons hanging without any justificatory picture.

Hume also argues that the fiction of personal identity is pragmatically and psychologically justified. First, the fiction of personal identity allows us to continue to speak justifiably about persons as they are normally construed. Parfit denied that we could do this because we had no metaphysical justification for the fiction of personal identity. As a result, under Parfit's picture we are denied the traditional psychological comfort of this fiction and are forced to rework our moral outlook to be properly reflective of some kind of correct metaphysical picture of persons. Contrarily, Hume has granted us freedom from metaphysical realist view of justification, and, in turn, the psychological comfort in keeping our traditional view of persons, as well as our traditional moral and prudential concerns. Second, Hume's account is

pragmatically justified. Allowing us to have a usable concept of persons enables us to have a functional government and legal system that can comfortably rely on the fiction of personal identity as a structural guide. As we have seen in Chapter Four, reworking the legal system to reflect a proper metaphysics results in wild indeterminacy and hopeless subjectivity, since it depends upon Parfit's speculative ontological claims about psychological connections as proper demarcations of selves. Ultimately, in addition to having a satisfactory explanation of how we arrive at the fiction of personal identity divorced from metaphysics, we can also justify this fiction *via* an appeal to pragmatism.

Now, we see that both Hume and Parfit believe that our beliefs about persons ought to be grounded in something. Johnston, on the other hand, just leaves these beliefs to be utterly basic and unproblematically ungrounded. So where is the value in bringing up Johnston's minimalist account? What *motivates* Johnston's argument that our beliefs about persons are utterly basic is his belief that the metaphysics of persons fail to provide a justification for this belief. And from the Humean stance, this is true, but it is *not because* a metaphysical picture of persons fails to justify our beliefs about persons that our beliefs about persons are reasonable.

This subtle difference between Hume and Johnston on the justificatory status of our person-directed beliefs demonstrates the difference between the Humean position and the Parfitian position most clearly. Parfit is a self-

described "realist" about importance.[99] This means that what matters are the metaphysical facts that constitute persons. On Parfit's view, the metaphysical facts constitute what matters, independent of the way we view or describe these facts. Korsgaard has interpreted Parfit's realist view as one committed to the idea that "normativity is an irreducible non-natural property that is independent of the human mind. That is to say, there are normative truths - truths about what we ought to do and to want, or about reasons for doing and wanting things."[100]

Parfit is committed to the idea that there are metaphysical truths that dictate whether our moral views are reasonable. What makes this a hard account to defend is that Parfit has the burden of proving how it is that metaphysical truths about persons entail certain normative properties. Hume, on the other hand, shows that normative truths are based on beliefs about persons, which are themselves justifiable. And these beliefs are justifiable because they, in turn, are based upon relations that we perceive and the conventions structured to reflect these relations.

The decisive difference, then, between Hume and Parfit is this. For Hume, the fiction of personal identity is justifiable because it is based on relations and conventions. And these relations and conventions come into existence via our interaction with the external world. They are useful to us,

---

[99] Parfit (2003): 308.
[100] Korsgaard (2003): 1.

and they structure the nature of our collective life. For Parfit, persons are simply reducible to the metaphysical truths that constitute them because these metaphysical truths bear, in and of themselves, what is important.

Parfit's account fails because it does not provide us an account of how we might be properly "hooked up" to the metaphysical facts so as to reflect the normative properties he has in mind. Hume, instead, shows us that our beliefs about persons are justified based on the relations we perceive and the conventions that arise to reflect them.

Conclusion

Parfit argues that while persons exist in some obscure sense, they do not exist in any sense that is irreducible to the facts that constitute them. Indeed, Parfit argues, the truth of persons "must consist in the truth of facts about bodies, and about various interrelated mental and physical events. If we knew these other facts, we would have all the empirical input that we need. If we understood the concept of a person, and had no false beliefs about what persons are, we would then know, or would be able to work out, the truth of any further claims about the existence or identity of persons. That is because such claims would not tell us more about reality."[101]

Hume's fictionalist account shows that Parfit's claim, that personal identity tells us no more about reality than its constitutive facts, is clearly

false. We can, indeed, make claims about persons that cannot be made true or false by a mere appeal to non-personal facts. This is because the fiction of personal identity is constituted not only by such facts. The fiction of personal identity is also constituted by, and made reasonable within, a network of conventions.

Ultimately, Parfit is left in the position of having to defend why it is that persons are simply reducible to non-personal facts when he admits that persons do, in some sense, exist. Parfit would defend his view by stating that he believes that normative properties consist in the bare metaphysical facts. He has defended this line when he claims that persons – or "conceptual facts"– are not important relative to the truth of the lower-level metaphysical facts.[102]

Parfit's account has two primary weak points. Parfit's metaphysical claims are problematic in that they do not, in fact, properly entail an ethical picture. Additionally, the indefensibility of his ethical theory is evidence against his metaphysical theory. As a "realist about importance," Parfit is committed to the dubious claim that metaphysical facts, independent of the mind, have moral importance. This, as Korsgaard points out, puts us "in a very small box."[103] The strength of Hume's account is that we can continue to speak about persons without making an appeal to metaphysics. At this point,

---

[101] Parfit (2003): 297-8.
[102] Parfit (2003): 308.
[103] Korsgaard (2003): 1

we can either accept Parfit's realist account of importance, wherein it is vague how persons come to regard these metaphysical normative properties, or we can accept Hume's account of persons, which doesn't rely on bare metaphysical speculations and which instead is justified via our relations and conventions. I argue that we must accept Hume's account of persons, because it shows how it is that we arrive at our person-directed beliefs, and how these beliefs are justifiable within the network of our conventions.

Ultimately, Parfit fails to show how a correct metaphysical view of persons, which dictates what is important, deflates our notion of personal identity and renders our moral practices unreasonable. On Parfit's account, the metaphysical idea of a Cartesian ego *could have* justified personal identity and our moral practices. But this metaphysical picture turned out to be false. When we illuminate the metaphysical truth, according to Parfit, we see that personal identity is not important. Hume decisively forges a link between our beliefs about personal identity and our moral practices, and the metaphysical facts of the matter. In this way, we do not have to be worried about getting the metaphysical facts right. Our beliefs about persons and our moral practices are justified without an appeal to metaphysics. And through Hume's approach, we can see that our person-directed beliefs, as they stand, are justified, and not tethered to metaphysical speculations about "real" normative properties.

Bibliography

Adams, Robert Merrihew. "Should Ethics Be More Impersonal?" in *Reading Parfit*, ed. John Dancy (Oxford: Blackwell, 1997).

Baier, Annette. A *Progress of Sentiments: Reflections on Hume's Treatise* (Cambridge: Harvard University Press, 1991).

Behrendt, Kathy. "The Neo-Kantian and Reductionist Debate," *Pacific Philosophical Quarterly* 84, 2003.

Dancy, Jonathan, ed. *Reading Parfit* (Malden, MA: Blackwell, 1997).

Dresser, Rebecca. "Life, Death, and Incompetent Patients: Conceptual Infirmities and Hidden Values in the Law," *Arizona Law Review* 28, 1986.

---. "Personal Identity and Punishment," *Boston University Law Review*, May 1990.

Eklund, Matti. "Personal Identity, Concerns and Indeterminacy," *The Monist*, October 2004.

---. "Personal Identity and Conceptual Incoherence," *Nous* 36(3), 2002.

Fogelin, Robert. *Hume's Skepticism in the Treatise of Human Nature*. (London: Routledge, 1985).

Hume, David. *A Treatise on Human Nature* (Oxford: Oxford Univ. Press, 1978).

Jeske, Diane. "Persons, Compensation, and Utilitarianism," *The Philosophical Review*, 102(4), October 1993.

Johnston, Mark. "Human Concerns without Superlative Selves," in *Personal Identity*, ed. Raymond Martin and John Barresi. (Massachusetts: Blackwell, 2003).

---. "Fission and the Facts," *Philosophical Perspectives*, 3: 1989.

---. "Reasons and Reductionism," *The Philosophical Review*, 101(3): July 1992.

Kant, Immanuel. *Groundwork for the Metaphysics of Morals*, ed. Thomas E. Hill (Oxford: Oxford University Press, 2003).

Korsgaard, Christine. *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996).

---. "Personal Identity and the Unity of Agency," in *Personal Identity*, ed. Raymond Martin and John Barresi. (Massachusetts: Blackwell, 2003).

---. "Normativity, Necessity, and the Synthetic a priori: A Response to Derek Parfit." This paper was written for a conference (on the moral philosophy of Derek Parfit) at Rutgers University in Spring 2003 and is available at Korsgaard's homepage:
http://www.people.fas.harvard.edu/~korsgaar/Korsgaard.on.Parfit.pdf.

Lewis, David. "Survival and Identity" in *The Identities of Persons*, ed. A. Rorty. (Berkeley: University of California Press, 1976).

Miller, Kristie and David Braddon-Mitchell. "How to be a Conventional Person," *The Monist*, October 2004.

Nagel, Thomas. "Brain Bisection and the Unity of Consciousness," in *Moral Questions* (Cambridge: Cambridge University Press, 1979).

---. *The Possibility of Altruism* (Princeton: Princeton University Press, 1979).

---. "Universality and the Reflective Self," in *The Sources of Normativity*, ed. Onora O'Neill (Cambridge: Cambridge University Press, 1996).

Noonan, Harold. *Personal Identity* (New York: Routledge, 1989).

Olson, Eric. *The Human Animal* (Oxford: Oxford University Press, 1997).

Parfit, Derek. *Reasons and Persons* (Oxford: Clarendon Press, 1984).

---. "The Unimportance of Identity," in *Personal Identity*, ed. Raymond Martin
and John Barresi. (Massachusetts: Blackwell, 2003).

---. "Normativity" (unpublished)

Perrett, Roy. "Personal Identity, Minimalism, and Madhyamaka," *Philosophy*

East and West 52(3): 2002.

Quine, W.V., review of *Identity and Individuation,* Milton Munitz, ed., in *The Journal of Philosophy*, 1972.

Rhoden, Nancy. "Limits of Legal Objectivity," *North Caroline Law Review* (July 1990): 854.

Sher, George. *Desert* (Princeton: Princeton University Press, 1987).

Shoemaker, David. "Theoretical Persons and Practical Agents," *Philosophy and Public Affairs* 25(4), 1996.

Shoemaker, Sydney. "Critical Notice of *Reasons and Persons*," *Mind* 93, 1984.

---. *Identity, Cause and Mind: Philosophical Essays* (New York: Oxford University Press, 2003).

---. "Persons and Their Pasts," *American Philosophical Quarterly* 7 (4), 1970.

---. "Self, Body, and Coincidence," *Proceedings of the Aristotelian Society, Special Volume* 73, 1999.

Unger, Peter. *Identity, Consciousness, and Value* (New York: Oxford University Press, 1990).

Williams, Bernard. *Problems of the Self* (Cambridge: Cambridge University Press, 1973).

Wittgenstein, Ludwig. *On Certainty* (New York: Harper & Row, 1969).

Wolf, Susan. "Self-Interest and Interest in Selves," *Ethics* 96(4), July 1986.