Thought Experiments in Personal Identity: A Literary Model

BY

ALEKSEI ZARNITSYN
B.A., University of Arkansas, 2002
M.A., University of Arkansas, 2004

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Philosophy
in the Graduate College of the
University of Illinois at Chicago, 2013

Chicago, Illinois

Defense Committee:

Marya Schechtman, Chair and Advisor
David Hilbert, Philosophy
Colin Klein, Philosophy
John Gibson, University of Louisville
David W. Shoemaker, Tulane University

To my parents.

Acknowledgments

The most thanks and gratitude goes to my thesis advisor, Marya Schechtman. Without her guidance, conversation, patience and encouragement over the years I would not have written this. Thanks also to Ed Minar, who got me started much longer ago. Thanks to my thesis committee for all the valuable guidance: Colin Klein, David Hilbert, John Gibson, and David Shoemaker. David Shoemaker's exhaustive, timely, and challenging comments went far beyond what one might expect from an external member of the committee. Thanks to Peter Hylton for his comments and guidance during the pre-dissertation stage. I talked to many people about this project, and have benefited from those conversations. Thanks to my friends and colleagues: Brian Casas, Bucky Farley, Bob Fischer, Jessica Gordon-Roth, Mae Liou, Sean Morris, Matthew Pianalto, David Schaffer, and to all my fellow graduate students in writing seminars, as well as my Russian friends with whom I discussed these thoughts in some rather unexpected settings. A very special thanks to a philosopher outside the academy, Irina Zarnitsyna. Finally, I want to thank my wife Mae Liou for her editing, and express my deep gratitude to her for being there through it all.

Contents

Summary

Thought experiments are standard methodology in the philosophy of personal identity. The cases used, however, are rather fantastic: fission, teleportation, fusion, and so on. Why should we think that what is essentially a fantastic story can aid us in delivering knowledge about the fundamental questions of personal identity? One might point to the success of thought experiments in science to support the use of the methodology in philosophy. However, there have been serious challenges to the suggestion that thought experiments used in personal identity can fit this model. On the basis of this critique, some philosophers have suggested giving up on this method altogether. I suggest a different option. In my dissertation I bring insights about what and how we learn from the imagined fantastic scenarios found in literary fictions and in bioethics to bear on our understanding of the use of thought experiments in discussions of personal identity. I argue that by employing these insights we can identify a legitimate role for fantastic cases in allowing us to unearth constraints on the intelligibility of lives we can envision ourselves to inhabit. Such cases thus provide important information about the general conditions under which persons can continue to exist. This calls for a change in our understanding of the function and the results of thought experiments, but this is a welcome and fruitful change.

Chapter One provides a detailed overview of the thesis.

Chapter Two describes a tension at the heart of the thought-experimental method. Either philosophical thought experiments fit the successful model from thought experiments in science, or they are mere fictions, and as such cannot give knowledge. I grant that there are serious problems with trying to avoid the first horn of this dilemma. However, I think that the second proposition is false, and it has not received enough attention.

In Chapter Three, I turn my attention to a discussion of the cognitive value of fiction in aesthetics. One proposal is to"instrumentalize" fiction as containing implicit arguments, hypotheses, and theses. This is problematic because we wanted to know how we can learn from fiction *as such*. There are powerfully defended non-instrumental approaches to the cognitive value of fiction. One such model may help us avoid the second horn of our

dilemma and bridge the gap between knowledge and fiction. Chapter Three gives us a general view of how we can learn from the specifically fictional elements in stories.

However, the lessons from Chapter Three may be too general for it to be obvious how they can help us achieve our ultimate goal: to show how philosophers of personal identity profit from the use of thought experiments. In Chapter Four, looking at the uses of imagination in bioethics, we gain an understanding of how fiction can help us with the more specific issue of the complicated mutual dependence between the embodied entity of a person and the practical concerns associated with it. These guided explorations give an example of constrained imaginative engagement.

Methodological lessons from Chapters Three and Four can be applied to the standard puzzle cases in the philosophy of personal identity. In Chapter Five, I propose that by *fictioning* thought experiments via continuing to tell the story of thought-experimental survivors and the background world in which the pictured transformations are presumed to be possible we get insights into the complicated relation between different aspects of our lives. Our judgments about such features are based on our intuitive ability to project ourselves and persons like ourselves into other possible worlds. While fallible, such judgments are not arbitrary, but rather based on the recognition of the different significance various features of our lives have for us.

There are serious objections to my methodological proposal. I have been assuming that the practical concerns revealed in further explorations of the thought-experimental background have direct bearing on the metaphysics of identity. One may object that questions of fact and questions of value (those that are explored by thinking about practical concerns) should be kept separate. In Chapter Six, I reply to one version of this objection.

The response given to the earlier objection suggests that there is some unified entity which is the locus of all of our practical concerns. In Chapter Seven its a separate chapter now? Just checking to make sure its not a mistake, this assumption is challenged by the idea that different practical concerns are grounded by different metaphysical relations,

and that there is no reason to seek a single grounding unifier for all of them. I argue that some notion of a unique grounding unifier is a conceptual presupposition of our discussion of practical concerns. As such, arguments for its elimination threaten the intelligibility of the entire discussion.

Chapter 1

Introduction

1.1 A Caricature

Suppose Assistant Professor Franz Kafka submitted his *Metamorphosis* to an esteemed committee of his colleagues, consisting of Derek Parfit, John Perry, Sydney Shoemaker, and some external readers, to be evaluated as a possible thought experiment to use in teaching metaphysics of personal identity courses. The first sentence of the submission reads: "As Gregor Samsa awoke one morning from uneasy dreams he found himself transformed in his bed into a gigantic insect."[1] The piece goes on for some several dozen pages, describing Samsa's life and eventual death after the transformation. Prof. Kafka claims that he was inspired by some of the scenarios authored by the members of the present committee, such as fission, fusion, and teleportation.

The committee decides that Prof. Kafka's residency in philosophy has been a mistake, suggesting a more suitable employment in the literary studies department. The office is promptly prepared on the 15th floor of the building, securing easy communication with the former colleagues via a side staircase.

The committee has been divided about the decision, but two reasons seem to be most prominent.

(1) Some members of the committee argue that Kafka's fiction does not fit the pa-

[1] In the original: "Als Gregor Samsa eines Morgens aus unruhigen Traeumen erwachte, fand er sich in seinem Bett zu einem ungeheuren Ungeziepfer verwandelt." (Franz Kafka, *Metamorphosis*, 67)

rameters of knowledge-seeking discourse and its constraints; storytelling is in a different "weight category" from science and argumentation. There is no argument, hypothesis, or anything of the sort in Kafka's fiction as such—it is not its purpose, after all—to give us knowledge. While the members of the committee use fantastic scenarios in their thought experiments, their function is different: philosophers use fantastic cases to conceptually separate the intricately intertwined aspects of our being, physical and psychological. The intuitions triggered by thought experiments are used as data that inform arguments for different accounts of personal identity.

(2) Some other members of the committee acknowledge the possibility of Kafka's work to illuminate some ethical and, more broadly, practical questions. Such details of Gregor Samsa's life, however, are not relevant to the questions that worry metaphysicians of personal identity who are interested in the persistence conditions of individuals who are persons, providing accounts of the transformations that would destroy such an entity, and so on. Presumably, Prof. Kafka was not exactly interested in such questions.

The committee wants to issue the following clarification about the possibility of incorporating Prof. Kafka's contribution, with suitable modifications, into the philosophical corpus. The committee recommends that attempts be made to appropriate Kafka's novella into the genre of thought experiments but only by mentioning the bare bones of the transformation when necessary. The *story* element of Kafka's novella (i.e., the fiction itself) in such appropriations should be ignored, and only the shortened presentation should be used as an immediate springboard for the discussion of philosophical theories of personal identity. Thus, once the case is presented as coherent, further details of Prof. Kafka's fantasy are only a distraction. As the committee sees it, the philosophical upshot of Kafka's work is this: that the story of the fantastic change undergone by Samsa is at all coherent and does not strike us as nonsensical is an illustration of the conceptual separation between bodily and psychological elements of our lives. Since Samsa's "mind" is coherently imagined to be "transferred" into a different body—so goes the philosophical appropriation—it is the psychological elements of our lives that get the upper hand in

the criterion of personal identity over time. I.e., what it is to continue to be the same person is for the person's memories, intentions, thoughts, plans, and so on, to continue into the future. This presentation of Kafka's case looks schematically like John Locke's classic example of the Prince whose soul "entered and informed" the body of the Cobbler. Locke used the case to argue that it is the continuity of consciousness, rather than that of substance, that secured personal identity. Thus, Kafka's case of Gregor Samsa can be used as intuitively supporting some psychologically based account of identity. While Prof. Kafka continues his residency in the literary studies department, the philosophy department recommends recruiting more majors by mentioning the exciting work in the context of their metaphysics courses, to draw some interdisciplinary connections.

I think this is a useful caricature of how literary fictions are treated in metaphysics of personal identity. In broad strokes, suitable for an introduction, we are given two options for approaching the issue. One: we cannot get knowledge from fiction because fiction just isn't in the business of delivering knowledge. Two: we cannot learn our metaphysics of personhood from fiction because fiction is great for discussions of emotions, practical concerns, ethical norms, and so on, but these issues have no bearing on metaphysics of identity.

This thesis is inspired by thinking that both options are questionable, and that there is much to be explored in the connection between literary fictions and philosophical thought experiments. Of course, the connection has been discussed, but usually the discussion proceeds—as witnessed by the clarification issued by the committee—to suggest some ways for philosophical *appropriation* of literary fictions as implicitly containing philosophically important propositions, as exploring some general themes of philosophical significance, as an illustrating some abstract thesis, or as containing an implicit argument. Thus, literary fictions may be incorporated into philosophy as performing the same role as the one assumed to be performed by thought experiments. There is a serious problem with this approach. Ironically, there are powerful criticisms of fantastic thought experi-

ments as being *mere* fictions—entertaining but not capable of living up to the standards of knowledge-bearing discourse. This verdict has to do with the obvious problems that plague many thought experiments: the scenarios are underdescribed; the constraints on armchair speculation are not very well articulated; our intuitions are subject to various biases, and so on. So, saying that literary fictions work like thought experiments inspires skepticism about both. Related to this, instrumentalizing fiction to serve philosophical purposes threatens to miss explaining how fiction itself can "ring true" and teach us—the powerful intuition that literary fiction as such can be cognitively significant.

In this thesis, I suggest that we should take seriously reversing the direction of fit in the obvious link between literary and thought-experimental fictions. Instead of thinking that in understanding the cognitive significance of literary fictions we are required to present them as being implicit philosophical thought experiments, we will benefit from thinking that understanding the cognitive value of fiction as such may serve as a useful model for understanding cognitive value of philosophical thought experiments in personal identity. The central move of the dissertation is thus the methodological reorientation of the purposes and aims to which we use thought experiments in the philosophy of personal identity, by looking more closely at the function of imagination in literary fictions.

This move is particularly welcome as moving forward the discussion of the function and value of thought experiments in personal identity in light of serious methodological misgivings about it. While the methodology of thought experiments is widely accepted, there is at the same time a certain tension at the heart of the method. As the fantastic cases to which philosophers appeal get stranger and stranger, it is not clear what separates the philosophical thought experiments from storytelling. One way to avoid this charge is to appeal to the analogy between the scientific and the philosophical uses of fanciful cases: they are used in science, and so they may be similarly useful in philosophy. This analogy fails, however. We are left, then, with the uncomfortable admission that the philosophical thought experiments are in fact more like fictional stories. At this point, one may worry whether anybody who is serious about the metaphysics of personal identity should bother

with such stories.

This worry arises, however, only under the assumption that fictions as such cannot be cognitively significant, and it is this assumption that I question in the thesis. This thesis articulates a literary model of philosophical thought experiments, provides a template for reinterpreting classic puzzle cases in personal identity, explains the costs of such a move (which may be significant), and replies to two serious theoretical objections against it. I want to be clear that I am not suggesting that we replace philosophical thought experiments with science fiction. Nor am I suggesting that the point of fiction is to "corrupt the youth" by smuggling a bit of philosophy into their texts, although that may be what always happens. Instead, my methodological proposal is to look at fantastic scenarios in their natural habitat, literary fictions, and to try to understand the specific resources offered by the uses of imagination in such contexts. In doing so, I aim to understand the possible contribution to our knowledge of similar sources of learning, when they are put to work in thought experiments in philosophy.[2]

Let me clarify this. It is useful, I think, to focus on the changes that my model suggests. First, there is a change in the question we are going to ask about our intuitive reactions to thought experiments. These reactions are standardly taken to offer support to this or that theory of personal identity. According to the standard treatment, the case of Gregor Samsa should help us address the question: "Does a person survive a radical change in her embodiment?" However, according to my model, the real question before us is whether the scenario described here—in the form of the thought experiment and the subsequent additions to it to clarify the thought-experimental background—is at all coherent, and whether asking this question can reveal other aspects of personhood. We can think of the general question I will ask about thought experiments along the following lines. Is the world described in the fiction here conceptually and materially similar (enough, roughly) to the world of our form of life? If we can make the case that it is, or that the imagined

[2]Thus, the elements of fictional works that make them candidates for the discussion of learning, that make the great-book approach standard in many institutions of higher education, may help us understand how we can learn from thought experiments in personal identity.

world gives us a coherent extension of our practices, then reflecting on the case can be helpful in understanding the interaction between the clusters of features that make our lives possible in such worlds as lives of persons. If our discussion of this question of coherence and consistency leads us to significant difficulties, or resolves in disagreement, then it is a sign that the case may not be able to help us address the question at hand, or, alternatively, that we may have to live with a disagreement in answers to some questions. Either way, we may gain a more nuanced understanding of the conditions under which persons can continue to exist.

I hope this is sufficient to distance my view from the idea that I am suggesting to read Kafka *instead* of Parfit, or to ask Parfit to write more and better fiction. I am suggesting that we should engage with philosophical thought experiments with these broader questions in mind and be prepared to pose questions about the scenarios that may have seemed out of place, or at best secondary, in the standard practice—for example, questions about the interaction between the purely theoretically conceived fantastic transformation and the practical contexts of our lives. In many cases, I suspect that we are not going to simply lift off answers to these questions from the philosophical scenarios as they are presented right now. So, the import of standard cases may have to be rethought, but part of the interest in engaging with thought experiments is in these further clarifications and explanations. At any rate this is not too much to ask, I think, since philosophical thought experiments do not exist in abstraction from other thoughts of their readers; imagining such cases is not independent from the practical contexts of our lives. It is no loss, then, to propose further *fictioning* of thought experiments by adding more details of the background world to the existing scenario.

So, roughly, additional details of such world-making (i.e., of elaborating thought experiments) are supposed to help us examine clusters of features of our lives that have to be presupposed when we discuss questions of personal identity. By imagining the practical context of the fantastic transformations, we envision possibilities of our form of life being able to incorporate the imagined change, and sustain it. We are not asking the

traditional metaphysical questions, such as whether a later person is the same person as the earlier one. Instead, we are probing the sustainability of our concepts in the newly envisioned circumstances that form the background of our explorations of the questions of identity. Such probings and explorations are defeasible, and I am not arguing that we can be certain that we are not simply being misguided by the thought-experimental fictions we craft. But I believe that if informed discussion of these cases can withstand informed criticism, this kind of coherentist justification is enough to get us going.

Here is the "skeleton" of the thesis, immediately followed by a more developed chapter-by-chapter summary.

1.2 The Bare-Bones Sketch of the Thesis

1. There is a tension at the heart of thought-experimental method. Either philosophical thought experiments fit the successful model from thought experiments in science, or they are mere fictions, and as such cannot give knowledge. I grant that there are serious problems with trying to grasp the first horn of this dilemma. However, I think that the second horn is false.

2. The second horn of the dilemma in (1) has received some (but not enough) attention. One proposal is to instrumentalize fiction as containing implicit arguments, hypotheses, theses. This is problematic because we wanted to know how we can learn from fiction *as such*. However, there are powerfully defended non-instrumental approaches to the cognitive value of fiction from aesthetics. One such model of knowledge in fiction may help us grasp the second horn of our dilemma in (1).

3. (2) gives us a general view of how we can learn from the specifically fictional elements in stories. But it may be too general to understand how philosophers of personal identity can profit from it—our ultimate goal. Looking at the uses of imagination in bioethical discussions, we gain an understanding of how fiction can help us with the more specific issue of the complicated mutual dependence between the embodied entity of a person and the practical concerns associated with it. These more narrowly guided explorations give

an example of constrained imaginative engagement.

4. Methodological lessons from (2) and (3) can be applied to the standard puzzle cases in the philosophy of personal identity. I propose that by *fictioning* thought experiments via continuing to tell the story of thought-experimental survivors and the background world in which the pictured transformations are presumed to be possible we get insights into the complicated relation between different aspects of our lives. Our judgments about such features are based on our intuitive ability to project ourselves and persons like ourselves into other possible worlds. While fallible, such judgments are not arbitrary. We can make justified (but fallible) predictions about what may happen to persons like us, based on recognition of the different significance different features of our lives have for us.

5. There are serious objections to my methodological proposal. (1)–(4) assume that the practical concerns revealed in further explorations of the thought-experimental background have direct bearing on metaphysics of identity. One may objetct that questions of fact and questions of value (those that are explored by thinking about practical concerns) should be kept separate. I argue that the separation in this case is unmotivated.

6. The response to the objection in (5) suggests that there is some unified entity which is the locus of all of our practical concerns. This is challenged by the idea that different practical concerns are grounded by different metaphysical relations, and that there is no reason to seek one grounding unifier for them. I argue that some notion of a unique grounding unifier is a conceptual presupposition of our discussion of practical concerns. As such, arguments for its elimination threaten the intelligibility of the entire discussion.

This concludes the skeleton of the general argument. Let me now go through the main points of my argument by giving a chapter-by-chapter summary of the thesis.

1.3 Chapter-by-Chapter Summary of the Thesis

Chapter Two presents the challenge for philosophical thought experiments. At least since the time of John Locke, (fanciful) thought experiments have been standard methodology in the philosophy of personal identity. But as the cases get stranger and stranger, it is

less and less clear how we should be able to learn anything from them.

Now, one common line of response to this objection is to appeal to the use of various hypothetical scenarios, including the fantastic ones, in science. How do such cases work in science? By carefully distinguishing between the background and the fantastic variable manipulated against that fixed background, we can perform controlled experiments in our heads, and try to determine which factors in the system are responsible for the varying outcomes when we tweak this or that variable. While something like this is presumed to be going on in philosophy, closer examination shows that the analogy breaks down. Suppose you take the teletransportation case to fit the scientific model of thought experiments. Let's assume that the background "theory" is a set of various assumptions about the behavior of persons. The variable is the idea of normal embodiment and some assumptions about the continuity of objects. Now suspend this assumption. In effect, you vary the embodiment–parameter, and you try to hold everything else constant: social interactions, laws, institutions, and so on. Run this experiment and collect intuitions about whether the person who exits the teletransporter is the same person as the one who entered it at the other end. For simplicity's sake, let's assume that there is some uniformity in our responses, such that we think that the operation preserves the original person. Again simplifying, the conclusion is that the continuity of the same functioning matter—in this case, the continuity of the person's brain—is not necessary for that person's continuation.

Kathleen Wilkes (1988) argued that this analysis is mistaken. In the case just mentioned, we assumed that we can hold the background fixed while manipulating the variable. According to Wilkes, however, our personhood terms do not function like independent variables in a tight system of theoretical assumptions. Instead, we should wonder whether in formulating the background we already presuppose, for example, certain features of our embodiment. But then suspending embodiment will mess up the background in ways we cannot exactly picture. Wilkes poses a dilemma: either thought experiments are impossible against the background as we know it; or if some other background is assumed, we don't know why speculations about that possible world should be of any

relevance to the question what we most fundamentally are. Her conclusion is to give up on thought experiments in favor of actual cases, which are in fact as challenging.

I don't question Wilkes's argument by directly addressing her complaints. I question her assumption that either thought experiments in personal identity fit the scientific model of success, or they are merely entertaining. Cora Diamond, among others, questioned this assumption, and I think we can use her insight to address Wilkes's concerns by pointing to a different way of thinking about the function of thought experiments.

The central move in this thesis is to shift our attention from "experiment" part of "thought experiment" to the "thought" part by acknowledging that philosophical thought experiments, properly understood, *are* like fictions, and *that* is the source of their perhaps unacknowledged potential. This move seems natural because philosophers are not the only ones to contemplate fantastic changes. Our cultural tradition is full of stories of doubles, species-boundaries transgressions, body-swaps, and other amazing transformations, and it seems that the existence of such a tradition expresses our deep preoccupation with change, identity, and mortality, and can point to ways in which we can reflect about ourselves using such transformations.

Chapter Three examines some current literature about the cognitive value of fiction, which seems like a natural move to pursue the central idea of the thesis. As I mentioned earlier, the points of contact between philosphical thought experiments and fiction have been discussed in the literature. What I call 'instrumentalist' readings of fiction attempt to incorporate literary fiction into the philosophical canon by seeing it as implicitly containing arguments, themes, or examples that can be used for philosophical purposes. Such instrumental assimilation has its problems, most centrally the loss of the fictional itself, which we aimed to retain in order to avoid Wilkes's criticisms. If we interpret literary fictions as implicitly containing arguments or philosophical puzzle cases, then Wilkes's worry applies to them as well. The examination of literary fictions, however, was supposed to show us how to avoid Wilkes's problem; i.e., it was supposed to show us new resources available in engagement with the fiction itself.

Turning to non-instrumentalist approaches to understanding cognitive value of fiction, I discuss in detail John Gibson's work. According to Gibson, literary fictions are cognitively significant as presenting us with visions of the embodiment and fulfillment of our conceptual knowledge in the axiological dimension of our lives. The function of such visions cannot be reduced to tools of conceptual clarification. Literary works put our concepts into the context of practical worlds of interpersonal interactions to reveal what we take to be valuable, significant, worthy. While literary fictions are dependent on prior knowledge of concepts, what they do is not cognitively trivial. Literary fiction is uniquely positioned to give clear and sustained attention to the aspects of human reality that are usually closed off from view due to the practical limitations of our lives, in which such sustained attention is hard to actualize.

At the end of the chapter, I review a sample of 'literary counterparts' of philosophical thought experiments: stories that bear some surface similarity to puzzle cases from the philosophy of personal identity. This brief examination is meant to point to ways in which fiction can be cognitively significant. I acknowledge that my readings may be deficient in numerous ways, but I hope that they serve their humble purpose.

Chapter Three gives us a general outline of the picture of cognitive value of fiction. There is a problem, however, with its overly general nature, and also with the difficulty of disentangling the artistic purposes of the author of the fiction from the cognitive value we are seeking. (Ironically, for example, the charge of latent instrumentalism can be applied to my model because of this feature of insufficiently contextualized discussion.) In Chapter Four, I explore thought experiments in bioethics to (a) narrow the general model from the previous chapter; (b) show that the fantastic scenarios we contemplate are on a continuum with actual cases and the cases of foreseeable future technological changes; and (c) provide an example of imaginative speculation that is guided and constrained. Because of the dynamic nature of our future-directed practices, contemplating further and further changes is a useful tool of reflection. In bioethics, we are explicitly concerned with the question of the interaction between our understanding of the embodied human entity

with the technological and metaphysical possibilities of its dynamic development, and a system of values associated with it. Changes in our ability to control and manage further and further aspects of our lives open up new negotiations about the shape of our value system. Further probing of our intuitive reactions to fantastic possibilities can identify clusters of features that can be ranked according to the importance we ascribe to them: certain features of our lives are easier to give up than others. Just as in the literary case, of course, such speculative explorations are fallible and may not in fact be accurate predictions of the future.

The lessons from Chapters Three and Four result in a literary model of thought experiments. The model requires that standard puzzle cases be subjected to further background-world construction, and in general further story-telling about the possible futures of thought-experimental survivors. While there are some similarities with Wilkes's model, there are clear differences. Even though we agree on the importance of the background details, the use to which such details are put is different, and so is the general procedure. There is no assumption that we can fix the world prior to engaging in the story-telling, and the purpose for which we are doing it is rather different. Instead of asking the question whether a particular transformation results in death or survival, or whether one isolated feature determines our intuitive reaction, our question now is whether the scenario we are imagining is coherent and intelligible overall, given what we think of persons in our form of life. Admittedly, the analysis will have to proceed case-by-case, and the results from one case may not apply to another. As an outcome of this process, certain patterns of dependence between various aspects of our lives will emerge. By discussing our reactions in the exercise of imaginative speculation, we are clarifying constraints on our understanding of our form of life. Chapter Five applies this model to two famous cases: the case of fission and the case of teletransportation, and supplies a different interpretation of these cases. The model applies to other cases as well, but I do not pursue their analysis in this thesis.

Chapters Two through Five assume that practical concerns and metaphysical questions

of identity are tightly intertwined. This assumption has been shared, until recently, by much of the mainstream discussion of personal identity. It is natural to think that our intuitions about responsibility, compensation, and so on, determine the answer to the question of persistence conditions of persons. Thus, the intuitive reaction that the carrier of my psychology is responsible for my past misdemeanors no matter what happened to the body—Locke's Prince and Cobbler case—supposedly tells us that the persistence conditions for entities like persons are to be given in psychological terms. However, it has been under attack recently. For example, the proposal to sharply separate metaphysics of identity from the discussion of practical concerns is voiced both from the metaphysical and the ethical sides. This proposal is a serious objection to my model, and we have already seen it in the idea that the discussion of the quality of life has to be separated from the discussion of the persistence conditions of personhood. If this kind of separation is achievable, then further speculations about the practical and contextual background of our lives is indeed guilty of confusing metaphysics with other domains of inquiry. (At this point, recall the second reason that our caricature committee gave in its verdict about Kafka's fiction.)

In Chapter Six, I discuss two particular examples of this approach which I will label 'divide and conquer.' Eric Olson argues that our intuitive reactions having to do with practical concerns do not track the relation of numerical identity. They track the relation of psychological continuity, instead. Now, philosophers assume that psychological continuity is essential for individuals like us, but that is due to a confusion of the overwhelming significance of practical concerns with tracking identity. According to Olson, 'person' is not a good substance kind, and even if it were, it is not as attractive of a candidate as human animal. The intuitions that drive the psychological approach are compatible with animalism, which is more plausible on independent grounds. On the other hand, Susan Wolf attacks the purported conflation from the ethical side, arguing that whether metaphysical truth has any bearing on ethics depends on what value we ascribe to metaphysical truth, and that this issue puts us into the domain of value,

rather than metaphysical, discussion. In response to the general objection, I use Marya Schechtman's distinction between the question of conditions of value and the direct value questions. While personal identity theorists in the mainstream may not be asking the question about our substance kind, she argues that their question is not simply a direct question about value either. Rather, adopting Olson's language, the relation of 'being the same person' is a precondition of asking any of the specific questions about ethics or practice. It is then at a different level than the ethical questions about right and wrong and so on. Resolving the question of the ontological status of 'persons' is beyond the scope of this thesis, but I note that there have been sustained responses to Olson's animalism in the literature. As long as we can establish that psychological continuity theorists have been asking, in Marya Schechtman's terms, a literal question of identity, we can avoid Olson's and Wolf's objections.

Chapter Six suggests that there is some underlying unity relation that is presupposed by our practical concerns. However, practical concerns are numerous, and it is hard to see how one type of metaphysical relation can be a grounding unifier for all of the practical concerns. Recently, David Shoemaker articulated the pluralist approach with respect to the issue of metaphysical grounding of our practical concerns: different practical concerns are (possibly differently) grounded by different metaphysical relations, and seeking unification may in fact obscure more fruitful particularistic approach. In Chapter Seven, I address this objection. Based on the distinction used to address the earlier objection from Chapter Six, I draw a distinction between 'internal' and 'external' questions about our practical concerns. The question of unifying locus of our concerns is a necessary condition on asking the questions about particular practical concerns. I agree with Shoemaker that neither numerical identity nor psychological continuity can secure the foundation, but I don't think this is the end of the story in the argument against unification.

Chapter 2

Philosophical Thought Experiments Between the Scylla of Science and the Charybdis of Mere Fiction

2.1 Introduction

In this chapter, I discuss Kathleen Wilkes's famous criticism of thought experiments in personal identity (1988). Many philosophers appeal to the success of thought experiments in science to justify their use in philosophy. According to Wilkes, the analogy between the two fails because thought experiments in personal identity, as contrasted to those in science, are underdescribed. According to Wilkes, the typically abstract scenarios presented in philosophy have to presuppose some background, against which they are imagined. We have two options. First, we assume the actual world. Second, we assume some other possible world. According to Wilkes, responsibly spelling out the background in both cases exposes the problematic nature of philosophical thought experiments. Wilkes concludes that since we cannot regiment thought experiments on the scientific model (or at least that the thought experiments suggested up to this point do not fit the model), they cannot be fruitful. Her suggestion is to turn our attention to actual puzzling cases.

While I agree with Wilkes's criticism of the applicability of the scientific model of thought experiments to the philosophy of personal identity, resolving the issue along

these lines is not my purpose here. Whether Wilkes's conclusions can be challenged or not, I question Wilkes's assumption that if thought experiments in personal identity do not fit the scientific model, they are merely entertaining stories and cannot have the cognitive value required for metaphysics of identity. At the end of the chapter, I look at a similar move suggested by Cora Diamond's assessment of Wilkes's criticism and her suggestion that some thought experiments can be thought of as explorations, which bypasses Wilkes's dilemma. I suggest, however, that this move does not let one escape Wilkes's general question: why should thinking about the fanciful cases be enlightening even as exploration, given that they are just stories we tell? In the following chapters, I begin to address this question by providing a model of the cognitive value of fictional discourse.

Section 2 provides the definition of 'thought experiment' I will work with. Section 3 details Wilkes's criticisms of the method. Section 4 is a discussion of attempted responses to Wilkes, with some suggestions of how one may proceed. Section 5 suggests that Wilkes's vision of what thought experiments are for and the possibilities of their productive use is rather restricted, and discusses Cora Diamond's proposal that some thought experiments are exploratory rather than decisive. But I argue that Wilkes's general point applies to this suggestion as well.

2.2 Thought Experiments

According to Wilkes, in a typical thought experiment, we (1) imagine a possible world "in which the possible state of affairs actually occurs—a world like our own in all relevant respects except for the existence in that world of the imagined phenomenon"; and (2) try to draw implications for "'what we would say if' that imagined set-up were to obtain; that is, if we inhabited that possible world" (1988, 2). While one may worry that Wilkes's definition is somewhat restrictive due to the language of "inhabiting the possible world" or looking at "what we would say" and therefore may exclude other ways of looking at and evaluating thought experiments, I think it is in line with more general definitions of

thought experiment, and where it is not, the difference is not damaging.[1] For example, Tamar Gendler defines thought experiment as follows. First, an imaginary scenario is described. Second, an argument is offered that attempts to give the correct evaluation of the scenario. Third, the evaluation of the imagined scenario is then taken to reveal something about cases beyond it (Gendler 2002, 21).[2]

The scope of Wilkes analysis is intended to cover thought experiments like Plato's *Ring of Gyges*, amoeba-style division of persons, fusion of persons, teletransportation, and others. In each of these cases, according to Wilkes, we are asked to imagine a possible world and then make a pronouncement, or an inference, concerning a particular feature about which we are puzzled. None of these scenarios, according to Wilkes, is realizable in the world as we know it. Wilkes calls those thought experiments 'impossible' (1988, 2). We will come to the discussion of the kind of possibility she has in mind in a moment, but let me just mention outright that her usage is idiosyncratic and will not map neatly on the accepted classifications of modal terms (see Vaidya 2007). The kind of possibility she is concerned with is what she calls the "theoretical possibility": something is theoretically possible as long as it is allowed by the current understanding of the scientific theories, broadly construed, including some rough understanding of the likely trajectory of the scientific development of a given science. Her analysis excludes the hypothetical thought experiments that could be, but aren't for various reasons, carried out in practice. For example, the so-called 'trolley problems' from ethics are hypothetical because they can be carried out. Wilkes's analysis also excludes thought experiments from linguistics, like introspecting sentences for grammaticality, which should fall into the category of real experiments with mental entities (1988, 3).

[1] Thanks to Colin Klein for this point.

[2] Other formulations similar to this can be found in Norton 1996, Sorensen 1992, Rescher 2005, Parfit 1984. It may be instructive to mention Parfit's thinking on the issue at this point, since he is going to figure prominently in the thesis. Parfit says that contemplating imaginary cases, we "discover what we believe to be involved in our own continued existence... we discover our beliefs about the nature of personal identity over time" (Parfit 1984, 200). These beliefs are not merely about the imagined situations, but presumably "cover actual cases, and our own lives" (Parfit 1984, 200). Imagination must bring to the surface what is usually hidden from sight: "our beliefs are revealed most clearly when we consider imaginary cases" (Parfit 1984, 200). Beliefs about personal identity that are revealed to us in thought experiments are supposed to be universal and general.

2.3 Problems with Applying the Scientific Model in Personal Identity

Often philosophers appeal to the success of scientific thought experiments as a *prima facie* justification for their use in philosophy. It is said that Einstein, Mach, Galileo, among others, have used thought experiments to advance important theoretical breakthroughs (e.g., Wilkes 1988, Sorensen 1992, Parfit 1984).[3] According to Wilkes, the analogy requires critical attention: we need to look at the reasons *why* (some) scientific thought experiments succeed, and then investigate whether the analogy is warranted. Wilkes ultimately wants to show that if philosophical thought experiments are worthy of their title, they have to be subjected to the same kind of stringent constraints we find in sciences. The problem, however, is that the main feature of scientific experiments, namely careful specification of the relevant background, against which this or that variable is manipulated, is neglected in the philosophy of personal identity. There are two main difficulties Wilkes identifies with philosophical thought experiments: the theoretic background is underdescribed, and the difficulty in our understanding the relation between intuition and possibility. I will discuss each in turn.

2.3.1 Relevant Background

Wilkes (1988) attributes the success of the best scientific thought experiments to the fact that we can successfully identify a chosen variable, separate it from the background, and manipulate it against this fixed background. As we will see, in philosophical thought experiments, the background is underdescribed. According to Wilkes's, once we try to spell out the relevant background, the scenarios turn out to be problematic.

Scientific Cases

Consider the following thought experiment from Stevinus. Suppose two frictionless planes joined at an angle, supported by the third plane. I.e., visualize a triangle. Stretch a chain

[3] But see Reiss 2002 for a criticism of this position.

over the two sides of the triangle pointing up, making the lengths of each part of the chain equal the lengths of the plane on top of which it is stretched. (So, imagine something like a necklace hung on a triangle.) Now let the chain go. Will it move, or will it remain at rest? Stevinus said that the answer was obvious once we imagined joining the two loose ends of the chain. Since these are frictionless planes, if the chain moves, then gravity should be pulling it *forever*. But this is impossible, given the assumption that there can be no perpetual motion machines. By reductio, we conclude that if we let the chain go, it will remain at rest. Furthermore, this fact must be explained by the distribution of forces on the chain on the inclined planes (Wilkes 1988, 4). This thought experiment, then, by postulating something that does not occur in nature—frictionless planes—gets us to a useful result about force distribution on inclined planes.

Now, we know that frictionlessness is just a stipulation. Even though it is not something we can expect to occur in the actual world, what allows us to make pronouncements about this case, according to Wilkes, are established rules about treating different forces as additive, assuming that frictionlessness is the limiting case of the gradual reduction of friction we can achieve in the lab (1988, 31).

In contrast, suppose for a moment that suspending friction compromised holding gravity constant and/or made its behavior unpredictable. In this case, we could not have learned much from the experiments, because we could not assume the central feature of the experiment against the *fixed* background. In this case, our postulate would be "destructively relevant" to the rest of the theoretical background: we could not rely on the rules of discussing the contribution of other forces to the overall behavior of the system while ignoring friction.

Consider another example. Einstein supposed that if a man could run along the beam of light at the speed of light. According to Maxwell's theory of electrodynamics, this observer should see a stationary oscillatory field, but this is not something that could exist. In this case, since the truth of electrodynamics does not depend on physical limitations of our bodies, imagining this impossible feat of endurance is not destructively

relevant to holding the rest of the relevant theoretical background fixed—biology is not in the picture here. The language of human observation here is just a convenient device for establishing the implications of Maxwell's theory, and the adoption of this heuristically convenient description does not obscure the results of the thought experiment (Wilkes 1988, 3–4).[4]

Evaluating both examples, we appeal to the established conventions about what kinds of features are relevant for what kinds of questions. In general, science embodies this kind of approach really well precisely because its business is to isolate various classes of phenomena, about which we can make generalizations in different (often artificial, to varying degree) sets of conditions. For example, describing particle physics, Wilkes writes: "we know under what conditions a particle may be affected by *only* electromagnetic forces, and we know what follows if these circumstances obtain"(Wilkes 1988, 14). As long as the postulated theoretical impossibility does not affect the applicability of our agreed-on conventions of making inferences or judgments, it is not destructively relevant to the conduct of the thought experiment.

In addition to the method just described, we can think of the following ways of postulating a theoretical impossibility that may give us useful results. On the one hand, it can show us that there is some inconsistency in our initial assumptions. [5] On the other hand, we can suppose thought experiments that simply manipulate a variable against some arbitrary background of arbitrarily fixed parameters. In both of these cases, Wilkes's point stands: to draw any determinate conclusions from these uses of thought experiments, we have to supply the "rest of the game"—the rest of the theoretic background which

[4]See Norton's papers for a deflationist account of thought experiments. According to Norton, the surface presentation of the thought experiments like this is not what is doing the real work.

[5]One might think that the value of some thought experiments in science lies in their ability to show serious difficulties in existing understanding of a given theoretical framework. Thus, Einstein's thought experiment above can be one such case, showing that Maxwell's electrodynamics is problematic. Similarly, one might say that Galileo's thought experiment with falling masses demonstrated that the Aristotelian physics led to contradictions. Isn't the postulate in both of these cases in some sense "destructively relevant" to the assumed background in that it compromises it? No. In such cases, the manipulated variable does not affect our ability to hold the background fixed. It shows some problem with the theoretical background that can be constructed. The issue will become clearer when we come to the discussion of personal identity cases.

must be presupposed for these conclusions to be of interest; otherwise, our conclusions are problematic and may not be taken seriously.

As an afterthought to this discussion, Wilkes includes a section on the difference between the scientific terms as contrasted with the terms of philosophers—"the rich and glorious chaos of common sense" (1988, 13). While the "natural kind" terms of science have tidier definitions, clearer implications, and aim at explaining phenomena captured by scientific generalizations, "the ascriptions of common-sense terms have conceptual *links, suggestions, mutual involvements*—but few clear *entailments*" (Wilkes 1988, 13). Since she is questioning the ability of philosophical thought experiments to fit the constraints imposed by the model from science, the contrast between the terms that figure in scientific generalizations and much philosophical work further emphasizes the problems with constructing genuine generalizations and conventions for fixing the background world. If we are working with common-sense terms, and if these terms have mutual involvements, and are characterized by suggestion more than by clear rules and entailment, then holding the variable separate from the background becomes much more problematic. All this compounds the difficulties of spelling out the relevant theoretic background and thus compromises the attempts to fit the philosophical thought experiments into the scientific model. Let's look at some cases from philosophy to illustrate the difficulty.

Personal Identity Cases

Our concern now is whether there are philosophical thought experiments that are analogous to those in science, in which the imagined impossibility is not destructively relevant to the fixed background, against which we manipulate some variable (Wilkes 1988, 9). If so, then the philosophical thought experiments that invoke what is currently theoretically impossible may succeed just like the best cases from science. In order to address this question, we need to look at Wilkes's remarks on what the point of philosophical reflection on personal identity is because it will have direct bearing on what details one considers relevant to the issue at hand.

According to Wilkes, the relevance of the information to be considered in our theories of personal identity depends on the fundamental presuppositions about the questions we expect it to handle. She considers the standard identification (what it takes to be a person) and reidentification (what makes person A the same as person B) questions of personal identity (Wilkes 1988, 21–22). Wilkes says that the officially stated purpose of thought experiments in personal identity is to reveal "the heart of our current, present, notions of what it is to be a person"(1988, 12). For Wilkes, this is directly connected to the discussion of our practical concerns, embodied in our institutions, practices, and so on, including science.

Notice that Wilkes explicitly assumes that practical considerations are directly relevant to metaphysical questions of identity. This assumption has recently been under attack from philosophers with very different sensibilities and agendas (e.g., Schechtman 1996, Olson 1997, Shoemaker 2007). In Chapters Six and Seven of this dissertation, I address this issue more explicitly, and here I can say a few words about it. Since not all philosophers share the assumption that the relation between identity and practical concerns is unproblematic, one may wonder whether we can avoid Wilkes's criticism by simply denying the assumption. According to the most radical version of this move, insisting on further articulation of the background, including the practical background, can simply be irrelevant to metaphysics of identity, on some views (e.g., Olson 1997). Consequently, Wilkes's argumentation is limited in scope since her argument covers only one possible way of relating practical concerns and metaphysics of identity. In response to this objection, I think it is fair to say that Wilkes's critique sufficiently compromises much of the mainstream writing on personal identity because the assumption is widely shared. Therefore, discussing Wilkes is a useful way into a general discussion. Of course, at the end of the day, a more refined understanding both of the questions of personal identity and the possible connection of metaphysics to practice may change the assessment of Wilkes's work. But this is one of the points I want to return to in later chapters.

Having put this objection to the side for the time being, let's return to Wilkes, and

the presumed connection between metaphysical questions of personal identity and the practical background. To apply the scientific model of thought experiments in the philosophy of personal identity, we need some rough understanding of the term 'person' and the background 'laws' of personhood. Wilkes's own list of conditions of personhood draws from a fairly standard range of issues we associate with it: persons are rational; they are subjects to intentional ascriptions; we must take a stance towards them as moral objects; they can reciprocate that stance; persons are language-users; persons are capable of a consciousness of second-order states (with a version of free will); and, possibly, personhood has to do with using tools (Dennett 1976; Wilkes 1988, 23).

Of course, we are not going to get anything exactly like scientific clarity and precision in this domain. However, keeping this in mind, we can treat conceptual, institutional, scientific, and so on, background as the fixed point, against which we can assess the influence of a particular fantastic manipulation of this or that feature we associate with personhood.[6] Consider a possible amoeba-like splitting of persons to address the question of whether a person can survive such an ordeal.[7] According to Wilkes, before we draw any conclusions from this exercise of imagination, we had better supply the background details of the world, to give some criteria of intelligibility of this question. She asks, about this kind of case:

> [H]ow often? Is it predictable? Or sometimes predictable and sometimes not, like dying? Can it be induced,or prevented? Just as obviously, the background society, against which we set the phenomenon, is now mysterious. Does it have institutions as marriage? How would that work? Or universities? It would

[6] One might wonder whether calling the background 'conceptual' makes it sound as if we are misrepresenting thought experiments to be just about our language and not about ontological entities like persons. Wilkes does not discuss the point explicitly. We can probably assume that for her the discussion cannot proceed separately for concepts used to pick out the phenomena associated with personhood and for persons as entities. Our language does not float free of the capacities and behaviors of actual persons, and so when we are asked "what we would say" about this or that thought experiment, we are asked whether our concepts can pick out things in the possible world that are recognizably person-like. And the way we pick out recognizable person-like entities depends on the kind of life is being led by these entities. If they do not—that is, if we have no idea what to say confronted with the scenarios—the failure is on both levels: we can't recognize the world to be the kind of world we are comfortable describing in familiar terms.

[7] The case is from Sydney Shoemaker and it is mentioned in Parfit. Parfit's fission case is a version of it.

> be difficult, to say the least, if universities double in size every few days, or weeks, or years. Are pregnant women debarred from splitting? The *entire* background here is incomprehensible. When we ask what *we* would say if this happened, who, now, are 'we'? (Wilkes 1988, 11)

This requires some unpacking. First, we have to be clear on whether splitting is something like an anomaly or whether it is a regular occurrence. Second, we have to be clear—and it may be different depending on the answer to the first question—what theoretical background is relevant to our understanding the impact of fission on our practices and institutions. Third, it is unclear what connection the world of regular fission would have to the actual world. According to Wilkes, in the case of amoeba-like division, or other cases philosophers may want to use, the criteria, by which we judge that a person survived as one of her offshoots (or both, or none) are tightly intertwined with the practical details of family structure, financial institutions, courts, and so on, in that fantastic world. Fixing this background (the humanities' version of the theoretical background in science) is what allows us to apply the concepts of personhood to pick out a recognizably similar individual, track her through time, and say things like: "This is recognizably (the same) person's life while this is not"; "This is a borderline case, of which we cannot judge"; or "He is not the same person any more." But the problem with the case above is that this background is not stated. So, it is not clear why we should trust our imaginations in such cases, because it is not even clear *what* it is that we are asked to imagine.

One may wonder, thought, why we cannot just say that this case is exactly like the case of the frictionless planes. We assume that the "personhood background" stays the same, and we simply tweak one of the parameters of the model. Recall that in Stevinus's case we did not raise questions about how often and when we assume frictionlessness, and we still drew some valuable conclusions. But then why bother with such questions here? Suppose I split tomorrow. Why wouldn't it show something important about personhood?[8]

I take it that the general pattern of Wilkes's criticism appeals to conceptual holism of our personhood-background, and so the case is different from the scientific cases, in which

[8]Why wouldn't it show, for instance, that identity is not what matters in survival (Parfit 1984)? I discuss this in later chapters.

we presuppose that the contribution of different forces used in some abstract model can be calculated separately. So, according to Wilkes, in the world as we know it, we have some set of criteria, by which we judge survival, prudential concern, sameness of persons, etc. This set is holistically related to and sustained by a set of social institutions and practices. The set is applicable to a range of observations that it is designed to practically handle. Wilkes's point is that if the kinds of transformations a human organism can undergo changes so dramatically so as to allow fission, fusion, and all sorts of other changes presumed in thought experiments, then the criteria by which we judge personhood, survival, sameness of personhood, and so on, are bound to change as well, following the change described by these transformations. But Wilkes argues that we cannot predict this mutual change without more background details. Without it, we cannot figure out whether the postulated variable is or is not destructively relevant to the conduct of the thought experiment.[9]

Now, one might object that Wilkes does not explicitly argue for the assumed holism. First, though, some argument for holism can be gleaned from her remarks on the difference between our commonsense terms and the "natural kinds terms." (See earlier) As you recall, commonsense terms philosophers resort to have mutual involvements rather than implications, and this seems plausible. Second, while Wilkes assumes holism, her objector assumes the analogy between the scientific cases and the philosophical cases. Since this is part of the issue at issue, neither of the parties can be said to have advantage with respect to this.

The force of this remark will become more evident when we consider why she thinks we can neither assume that the background can simply be assumed to be our world, nor think that it is helpful to stipulate a different possible world. I explain these matters in the next section, in which we turn to the second problem with thought experiments in personal philosophy.

[9]For a similar type of response, see Johnston 1992.

2.3.2 Intuition and Possibility

When discussing our reactions to this or that stipulation, we have to rely on some agreement about what is and what is not relevant to the proper assessment of the scenario. Roughly, we should have some agreement in intuitions about possibilities.[10] We can think of different kinds of possibilities in terms of constraints on the scope of the discussion of possible worlds: what possible worlds are we excluding from consideration by using this or that variety of possibility? Thus, logical possibility covers all worlds that are logically possible; physical possibility covers what is possible given a background of physical theories A, B, C; epistemic possibility covers what is possible given a subject's background knowledge, etc. As I mentioned earlier, Wilkes is interested in what she calls "theoretical possibility." She defines it as possibility given current scientific understanding of physics, biology, etc., and some extrapolation from this knowledge (what is more likely, given what we already know). According to Wilkes, it seems clear that thought experiments in science are after some brand of possibility different from logical possibility because logical possibility does not allow any scientifically interesting predictions.[11]

A thought experiment is conducted to 'establish a phenomenon' in thought, on the basis of which we can draw inferences. How do we determine, in scientific thought experiments, whether something is, or is not, theoretically possible? To provide a model of determining theoretical impossibility, Wilkes works through an example of determining whether it is theoretically possible that iron bars float on water. The backing theory of metals is relevant to the question, so the adequate specification of the possible world has to include facts about floating, metals, water, etc. So, given the science of metals and of floating, and assuming various things about the surface tension of water, the question

[10] The matters with discussion of possibility are not so straightforward, of course, but this needn't deter us. See the introduction to Gendler and Hawthorne 2002.

[11] Gendler and Hawthorne 2002 define the notion of nomological possibility thus: P is nomologically possible with respect to a certain kind of nomos just in case "P is consistent with the body of truths expressed by those laws"(2002, 4). They go on to say that conceivability is better suited to exploring what they call the metaphysical possibility "how things might have been',' which is supposed to be a primitive notion. Wilkes's theoretical possibility is a subset of nomological possibility, I take it, restricting the class of all possible configurations of laws of nature to the theories of modern science. For our purposes, we can continue using Wilkes's terminology.

about floating iron bars is this: can an object with the specific gravity within the range from 7.3 to 7.8 also have a specific gravity of less than 1 (Wilkes 1988, 18)?[12] Wilkes writes: "[o]nce we describe the situation adequately, once we hand ourselves the backing theory of metals, then the (relevant) impossibilities appear—indeed, in the case of the iron bars the theory turns the impossibility into a logical one: nothing can both have an not have a specific gravity of less than one" (1988, 18). So we can say that iron bars cannot float on water in any possible world with the same theory of metals, same theory of water, same theory of floating, same understanding of 'bar,' and so on; otherwise, we get a contradiction. This illustrates the difference between *picturing* a floating iron bar and imagining the relevant background for the conduct of a scientific thought experiment about floating iron bars. For Wilkes, theoretical possibility is the common currency used to generate agreement in intuitive judgment. Put stronger, according to Wilkes, the shared intuitions in the scientific cases "seem rather to be straightforward inductive or deductive inferences" (1988, 15).[13]

What if we insisted that for the purposes of the thought experiment we should adjust the theory to allow the metals to float? Wilkes argues that the move is costly because of the adjustments we would have to make in other theories tied to the one we are working with. Changes in chemical laws would require changes in physical laws, and so on. Wilkes argues, in effect, that changing certain theoretical parameters roughly *here and there* cannot be done in good faith: you have to change a lot more and possibly everything, to supply the relevant background of the intended accommodation. When all the adjustments to our theoretical background are made, we effectively end up with a completely new theoretical background (Wilkes 1988, 30).[14] This in itself is not a problem. But there may be a problem with drawing conclusions about our world by looking at a

[12]Objects that float on water in normal conditions have specific gravity less than one, while metals do not meet this requirement.

[13]John Norton has a deflationary account of scientific thought experiments, according to which they are just dressed-up arguments, and their powers are to be explained by explaining the success of inductive and deductive inferences. This is in agreement with Wilkes's assessment: the talk of intuitions here disguises what is doing the work—our shared assumptions, from which the conclusions of thought experiments follow. See Norton 1996, 2004.

[14]This is obviously related to the discussion of the passage from Wilkes I quoted in full earlier.

(significantly) different one. Even if we are able to construct a possible world and trace implications of certain facts according to the rules of that world, why would observing changes of different variables in some world that is theoretically significantly different from ours tell us anything about our world? Furthermore, we might even grant that that is theoretically useful for understanding the shape of our theoretical knowledge—metatheory, as it were. Still, Wilkes would insist that this preserves her main point that to get at any results, relevant or not, we have to rely on fixed theoretical background. And, finally, even if this is useful for sciences, when we come to issues of personal identity, the case for this will be much harder to make.

Let's see how Wilkes applies this line of reasoning to some standard thought experiments in personal identity: brain bisection, memory-swap, and atom-by-atom recreation of a copy of an individual. In each of them, Wilkes argues that what is established is not a genuine theoretical possibility, and so the conclusions drawn from them by the users of thought experiments are not warranted. In her discussion, Wilkes draws heavily on our current scientific understanding of the relevant issues to question the theoretical possibility of the standard thought experiments. I think this shift of argumentation may obscure the philosophically stronger point discussed in the previous section. I discuss this after I go through the examples.

Take brain bisection. Many of such cases follow the following pattern.[15] Start with the documented actual cases of survival with one hemisphere when the other one is removed. Now, suppose your hemispheres are fully redundant: that is, all you need is one. Suppose that instead of discarding one hemisphere, as in the single case, its functionality is realized in some different body, your twin's. For symmetry's sake, also suppose that both of your hemispheres are implanted in the bodies of your twins. Everything that is necessary for survival in the single case is still there in the double case because in such a double transfer case, your relation to each hemisphere is the same. Both of the inheritors have your memories, desires, intentions, and so on, but they certainly cannot be identical to

[15] In Chapter Five, I will discuss Parfit's version of the case.

you since two things cannot be identical to one. The purpose of the thought experiment—at least for Parfit—is to sever the connection between our concern for the future and the relation of identity. The concern for each of the inheritors of my psychology seem to be rationally required despite the failure of identity.

Here is Wilkes's treatment of the case. Establishing theoretical possibility of brain bisection has to be guided by what we know from the sciences because it is part of "the backing theory of the human species." Wilkes's most forceful points rely on the role that the sub-cortical regions and the spine play in brain activity that is implicated even in higher order functioning that constitutes the production of phenomena we normally associate with unique personhood. To assume that the whole brain division preserves the functionality of each half, we have to make at least two very problematic assumptions: that the hemispheres are equipotent and that the division of sub-cortical regions and possibly the spine preserves the functionality of each resultant half. According to Wilkes, we have no reason to think that dividing the lower brain can be performed without destroying or significantly modifying the functionality of the resultant product. "Thus, the thought-experimenter has to suppose that all the non-cortical regions can be surgically sectioned too; *or* that the bits that cannot be divided are not essential to an individual's adequate, particular, and idiosyncratic modes of intellectual operation." Moreover, as Wilkes says, "there are good reasons to suppose that subcortical structures are connected both ipsilaterally and contralaterally thus indifferently" (1988, 38–39). The upshot is that even if the physical division of the brain can be performed, we have no evidence to think that the result will give us two functioning psychological continuers, needless to say two psychologically similar competitors, as many thought experiments in philosophy assume. But then the thought experiments, as they are presented, cannot really serve the purposes to which they are called.

Wilkes admits that the brain bisection may in fact be theoretically possible, but the general argument against the usefulness of this thought experiment for the officially stated purposes still stands. The thought experiment was supposed to provide us with enough

background in order to make inferences about the question of our interest—personhood—that are as little questionable as possible. However, there are serious difficulties with respect to the central supposition of the experiment: that the presumed division of the brain will result in the preservation of the functions relevant to personhood, twice over. Now, according to Wilkes, even if it is not conclusively theoretically impossible to divide the brain, in the face of serious doubts that it is in fact possible, the conclusions we draw from this are surely very questionable. At the very least, then, we cannot rest our theories of identity on such doubtful foundations.

Now consider memory-swaps that figure in some of the thought experiments. John Locke's classic case of the Prince and the Cobbler is something like a memory-swap case.Due to both the structural and the functional plasticity of the brain, simplified descriptions of memory-swaps are also subject to criticism. According to Wilkes, "if we are going to "rewrite" anything at all, given the holistic nature of the mental—substitute whatever it is that the corresponding notion is for the neural machinery of the brain to express the idea of 'holistic'—we have to rewrite the entire brain" (1988, 38–40). Wilkes admits that even if we can get pretty reliable results of recreating some *capacities* of one brain in another, the insurmountable problem is that "we will be unable to isolate the individual contents of individual mental states (beliefs that p, desires for x, expectations that q, and so forth)" that will secure something even close to similar mapping of one adult brain onto another (1988, 42).

What about creating an exact duplicate of a brain from scratch? Wilkes thinks that in order to get the exact duplicate, the rewriting must happen instantaneously, which is not theoretically possible (1988, 40, 42). If it is not instantaneous, however, we do not get the result we wanted.

Now, again, in these two cases, it may be arguable that Wilkes has not established the impossibility of such things as described by the users of thought experiments. So, one may say that unless she conclusively shows that such cases are theoretically impossible, these are legitimate possibilities to be considered by philosophers. I don't think that this

is a fair move, for the following reason. Wilkes is at pains to give some concrete proposals with respect to what counts as a theoretical possibility. If we agree that something has to be said about the theoretical background of the scenarios used in philosophy, then both parties have to present some reasons to think that the scenarios described in thought experiments have some connection to what we know. I think Wilkes's main point here is that appeals to ignorance with respect to future changes in our theories, or with respect to our ability to conclusively show that something is impossible are weaker arguments than those that rely on concrete evidence. On the weakest version of Wilkes's conclusion, the burden is still on the users of thought experiments to provide sufficient motivation to take seriously the fantastic cases.

To put this differently, appeals to ignorance described above begin to look like appeals to logical possibility of the thought-experimental scenarios. However, if we use this grade of possibility, then, according to Wilkes, it is not clear why the very same philosophers discussing these fanciful stories focus on the brain rather than directly discussing fictional stories, which may be describing logical possibilities. So, we have to stick with the theoretical possibility, in which case we cannot appeal to ignorance when constructing thought experiments (Wilkes 1988, 44-45).

In sum, Wilkes (1988) concludes with a dilemma:

> [E]ither (a) we picture them against the world as we know it; or (b) we picture them against some quite different background. If we choose the first, then we picture them against a background that deems them impossible... If we choose (b), then we have the realm of fantasy, and fantasy is fine to read; but it does not allow for philosophical conclusions to be drawn, because in a world indeterminately different we do not know what we would want to say about anything. (46)

Earlier, I mentioned that there seems to be an argumentative shift (perhaps a shift of emphasis) in Wilkes's reasoning. Discussing brain fission, memory-swap, and atom-by-

atom recreation, she appeals to contemporary science to question the assumption that these cases are theoretically possible. What seems to have dropped out of the picture is the discussion of the "heart of our current understanding of persons"—the practical and institutional background of the discussions of personhood. So, suppose that science undergoes a change that would make the kinds of transformations Wilkes envisions possible, or suppose we can fill in the missing details. Let's grant that if we put the emphasis on science, we lose the argument that the fantastic scenarios are impossible, and with it we lose part of Wilkes's criticism. This is not a critical loss, however. I suggest that we see the second stage of her argumentation, namely, the appeal to current scientific understanding of brains and such as a further reminder that all such discussions have to be placed in rather constrained context of theoretical possibility. As such, this point should at least make suspect our trust in these arguments without further details. Whatever is the case with this aspect of her argument, however, the domain of our discussion places us in the institutional and practical setting. Wilkes argument about inadequate background specification (of the institutions and practices) is not simply at the mercy of whether she is right on science.

2.4 Responses to Wilkes

2.4.1 Relevant Background Specification

The first difficulty with philosophical thought experiments, according to Wilkes, was insufficient specification of the relevant background, in the absence of which the inferences from the intuitions about what would happen in the contemplated world to a theory that accommodates these intuitions are problematic. There are a couple of replies one can advance against this idea. First, by itself, the fact that in the standard thought experiments the background is not specified does not amount to an 'in principle' objection. Second, as Wilkes claims, decisions about what background is relevant depend on the kind of question (of personal identity) that we are tackling. One might say that perhaps she may have misidentified the question or neglected the alternatives. It could be that

answering some of these other questions of personal identity may not require such detailed specification, after all.[16]

Soeren Haagqvist (1996) formulates the following response to Wilkes: the objection that the relevant background is not adequately specified cannot rule out thought experiments on principled grounds because it does not establish that future thought experiments won't succeed in spelling out all the necessary background details to the skeptic's satisfaction. In some cases, philosophers have been sensitive to presenting more refined thought experiments.[17] Similarly, Snowdon (1991) notes that the details can be supplied in later discussion. If such details are available later, then it would not be wise to prohibit the practice of entertaining thought experiments at the outset.

This line of response seems weak: one has to give reasons to think that such specification is in fact forthcoming before appealing to it. But granting this objection, Wilkes's argument may be weakened to say that up to now no thought experiment has done the job of adequate specification. Then the intuitions generated by these thought experiments that up to now that have been driving the debate can only be accepted as useful on the condition that further specification is forthcoming. If Wilkes is right, we cannot be sure that there has been a good thought experiment so far, which is not the conclusion many would welcome. So, even though Wilkes may not have ruled out thought experiments as fruitless across the board, the argument is damaging enough. Haagqvist thinks that Wilkes's conclusion should be weaker: we may use thought experiments, but with caution (1996, 28). Wilkes does not disagree with that, I think, but argues that proper caution

[16]The questions from personal identity Wilkes considers are that of synchronic and diachronic identity: what it is that makes an entity at time t a person, and what it is that makes the person p2 at time t2 the same thing as the person p1 at time t1. These are the standard philosophical questions, but they do not comprise an exhaustive list (Shoemaker 2007, Schechtman 1996, Rorty 1976). Furthermore, as I said earlier, Wilkes assumes that all discussions of thought experiments in personal identity will have to operate against the background of social institutions, ethical norms, and so on. More recent literature on personal identity questions this assumption (e.g., Shechtman 1996, Shoemaker 2007).

I discuss this objection and its implications in Chapters Six and Seven. However, even though Wilkes's views may be faulted on these grounds, her criticism can be restricted to those who assumes—as she does—that there is a direct connection between personal identity and practical concerns that we appeal to in our imagination. Against those who assume such a connection (and there are many philosophers in this camp), I think that the criticism still stands.

[17]See Martin 1991, for example.

should in fact rule all of them out.

Snowdon (1991) argues that in many cases we can grant that the thought-experimental situation is possible without knowing how it came about. Not knowing in full detail how this or that came about does not single out situations that are imaginary from those that are actual. But then the problem is not that the situations described in thought experiments are fantastic. Is this reply sufficient to compromise Wilkes's position?

It is clear that there are actual phenomena for which we lack explanations: take examples of dissociative identity disorder or fugue states, or in general take whatever the problems the frontier of any science tries to explain.[18] Such phenomena challenge our scientific generalizations. In some cases, we restrict the theory to exclude these occurrences from the range of phenomena that this particular theory explains. In others, we modify the theory so that it can explain the phenomena. In yet other cases, we do not know what to say. But in all these instances, we are dealing with exceptional cases against the background of some accepted ideas of what may count as an explanation of this occurrence.

To address Snowdon's objection, we can point out that there is still an asymmetry between actual and imagined unexplained cases. Wilkes's point was to discuss the conditions on the success of scientific thought experiments. In actual occurrences we cannot explain, the world of actual phenomena and the current theoretical understanding can help us out, or not. The special reason we may find imagination not as trustworthy is just that: we rarely have the whole world to help us out, and we cannot presume that we can assume the familiar background. So, the asymmetry is that in the imagined cases, the pattern of explanation is disjunctive. Either the imagined phenomenon is possible against the familiar background of this world, or it is actually so fantastic that it is destructively relevant to the background of this world, and so it must be pictured against the background of some other possible world. Snowdon may reply that more background details can be supplied at a later stage, but then either it is going to be our world, or some

[18]This is a complicated question that may have to do with various levels of explanation, explanatory purposes, etc., and I won't go into those details that are not relevant to the general point.

other world. At any rate, there is an asymmetry between actual and imagined cases, even though it is not because the imagined cases are singled out simply on the basis of lacking explanation. Moreover, this discussion shows that we often need to be very careful about trusting what we say in the cases for which we lack explanation, actual or otherwise. Snowdon's reduction of the issue to lack of explanation does not address this point.

2.4.2 Vagueness of Commonsense Terms

Haagqvist comments that Wilkes overlooks the possibility that the vagueness of our commonsense terms of personhood can be remedied by making them more precise. It is true that some philosophical problems may be addressed by clarifying and sharpening our definitions. Presumably, the modern understanding of infinity may show that Zeno's paradoxes trade on the wrong understanding of the nature of motion. Can we similarly deal with the problem of personal identity—will making our terms more precise remove the worries associated with vagueness? There seem to be two ways to go about it. First, we can stipulate one of the definitions of 'person' as the best one for our purposes—whatever they are—and claim that deviant intuitions and understandings are based on using a different, less restricted, concept. The proposal might help us to avoid some problems, but leaves it unclear how to connect what we ordinarily think of persons to that new cleaned-up version. Remember that one of the starting motivations—at least for Wilkes—was to get at the heart of our current understanding of persons, which is admittedly fairly messy. We can of course claim that it is not the purpose of a philosophical understanding of persons to account for the mess of human language as long as it serves some useful philosophical purpose, which we define by stipulating the best definition of 'person.' Notice, however, that this is not the move that those whom Wilkes criticizes are making. All parties to this debate are presumed to want to capture as much of our shared intuitive understanding of persons, and legislative moves seem to go against the pressure to provide a comprehensive account. (I don't take a position on this issue, and I return to it in Chapters Six and Seven.)

Another proposal is to distinguish between different questions of personal identity: reidentification and characterization (Schechtman 1996),[19] for example, or animal identity and narrative identity (DeGrazia 2005), and so on, and then argue that we gain clarity about personhood if we approach these questions first independently and then at a later stage attempt some way of bringing them together. This latter proposal, unlike the earlier one, takes the multiplicity of our understandings of persons seriously and attempts to confront the difficulty of sorting out various components of these understandings that make the presumed monolithic treatment of the term either problematic or not as straightforward as it may have seemed.

Still, even if this proposal to separate different questions of personal identity that have been pursued confusedly is on the right track, Wilkes's main point seems to apply here as well. Her point about relevant background specification as a requirement on the conclusiveness of thought experiments should withstand even such localization of questions. I.e., the problem of providing the relevant background simply becomes a more specific problem for each of the questions of personal identity. (Again, I am not taking the stance on this general theoretical issue, leaving it open whether it can be resolved.)

In sum, then, Wilkes's arguments about insufficiently specified background do not seem to have been refuted conclusively. Instead, we see that her position can either meet these criticisms by suitable weakening, or the replies are so far in terms of promissory notes that need further spelling out to provide an adequate response to Wilkes.

2.4.3 Possibility

The second difficulty Wilkes examines is that spelling out all the relevant background conditions requires paying attention to the background scientific theories to establish the theoretical possibility of this or that thought-experimental phenomena. This issue is

[19] The characterization question asks whether a particular characteristic is attributable to a given person (Schechtman 1996, 76). The main difference between the reidentification question and the characterization question is that their relata are different: the reidentification question asks for a relation between persons or person-stages, while the characterization question asks about the relation between a person and various actions, experiences, and so on.

closely related to the one we just discussed: we can only say that something is theoretically possible if we have taken care to spell out the details of the background theories; otherwise, we are left with appeals to ignorance.

Haagqvist suggests that Wilkes's objections are not relevant to the purpose of thought experiments, namely, "to investigate 'conceptual' theses by studying counterfactual situations." He claims, for example, that "it seems that this thesis [that teletransportation, if possible, could be used to send persons to Mars], if true, may still tell us something interesting about the concept of personal identity" (1996, 30). But Wilkes would insist that her point is precisely that until we have established a theoretical possibility here, we cannot judge whether it is true, what one would say, and whether what we would say would in fact tell us "anything interesting" for the purpose of developing a theory of personal identity. As I discussed earlier, this kind of objection looks like an appeal to bare logical possibility, which is not satisfying.

Haagqvist responds that we have examples of impossible suppositions—like frictionless planes—that are clearly fruitful in science. According to him, if Wilkes is right, then even these experiments cannot be fruitful. But surely this is false. So, Wilkes is wrong. There are two responses available to Wilkes. First, in the case of frictionless planes, we do have a reliable model of calculating the contribution of other forces to the behavior of an object and tracing the implications of suspending friction; i.e., we have established conventions of talking about it even though there may be questions about exactly what frictionlessness is. The impossibility introduced by frictionless planes can be isolated from the background of other forces we can hold fixed. In contrast, it is not clear whether we can isolate, say, the bodily and the psychological components of a given human being and simply check what we would say when we interrupt the continuity of matter while preserving the psychology, or vice versa.[20] Even if such a move were legitimate, the problem, as we saw in the discussion of adequate background specification, is that we have no idea about the practical implications of this possibility, and so we would not know 'what to say' in this

[20]Sorabji (2006) presents teletransportation as a modern version of the Aristotelian and Stoic form–matter distinction.

case.

What makes science different is that we have a host of experiments which speak in favor of the idealizing treatment in the case of frictionless planes. We certainly have no experimental analogue for teletransportation and the like.[21]

Haagqvist provides another argument to the effect that Wilkes has not spelled out what exactly counts as a "theoretical possibility," which leads her to too stringent of a constraint on what is useful for contemplation. Even though he ultimately thinks that there is no neat taxonomy of modality, he claims it would be wrong to require that all thought experiments fit one sense of "possible" (Haagqvist 1996, 144).

Some may find this remark useful, but a look at Haagqvist's own proposal would not have made Wilkes unhappy. Haagqvist proposes the following necessary condition on a successful thought experiment "A thought experiment is successful only if there is a best accommodation S of (C, F1 and... and Fn) and S contains ∼W," where C is a counterfactual supposition, F1, ..., Fn are further assumptions that that are relevant for assessing C's possibilities, S is the set of statements that are consistent with C and F1...Fn, and finally W is some statement supposed to be false under C, but which is supposed to be true, given our theoretical background. Moreover, the more conservative the set, the better accommodation it is, where conservativeness has to do with the set's resemblance to the set of the actual world (Haagqvist 1996, 151).

In other words, Haagqvist agrees with Wilkes that it is not enough that we can state the counterfactual itself; we also need to take care to spell out various other features of the world—the accommodation set—that make it clear what theoretical commitments the truth of the counterfactual implies. It looks like now we have an improvement over Wilkes's view since it does not rule out thought experiments, but maintains a more rigorous criterion for their success. However, even though Haagqvist allows for careful use of thought experiments, he thinks that none of the standard thought experiments he selected as paradigmatic—*Twin Earth*, *Chinese Room*, *Brain in a Vat*, among others—is

[21] Reiss 2002 is instructive in connecting thought experiments in science to experimental confirmation.

conclusive, once we work out the accommodation. I suspect that it is even worse with the standard puzzle cases in philosophy of personal identity, but I won't attempt to apply his analysis to them here. What is important for my purposes is that at the end of the day Haagqvist's theoretical relaxing of the restriction would not legitimate the standard cases in personal identity. All thought experiments up to now—it seems—are failing the stated conditions of success. So, even thought Haagqvist thinks that his arguments show that Wilkes does not succeed in ruling out thought experiments in principle, his conclusion is in practice as skeptical as hers.

2.5 Back to the Starting Assumptions

In the preceding sections, we discussed Wilkes's arguments, along with objections and replies. Even though we have seen that Wilkes's position is not without some weaknesses, a defender of thought experiments has a serious challenge to deal with. Wilkes herself concludes that standard thought experiments cannot teach us anything about personal identity, and suggests that we turn to actual cases, which are in fact as puzzling as any a philosopher's armchair might have produced.

Wilkes's diagnosis of the problem and this sweeping conclusion that thought experiments are not fruitful for metaphysics of personal identity should give us pause. Despite the serious methodological problem, thought experiments are part of established philosophical practice. For one thing, a lot of interesting developments of general kind have come from contemplating thought experiments, even though they are not methodologically beyond dispute. One can disregard such a pragmatic justification and maintain that unless we have a clear understanding of the method, we should question its results. This would be too strong, I think. In addition, the pragmatic success of thought experiments in moving the discussion forward may also suggest that explicit alignment of philosophical thought experiments with the scientific ones, along with the model of a variable manipulated against the fixed background, disguises something else that may be going on in thought experiments in personal identity. I think Wilkes is correct in her criticism of the

scientific model of thought experiments in personal identity; nevertheless, a full examination of the role of thought experiments in personal identity has to question whether this assumption captures their status and whether it adequately presents the source of the insights that they may offer.[22]

To do this, let me start by looking at Wilkes's own remarks about some philosophical thought experiments and literary fiction, because she draws a non-accidental and telling parallel between them. Looking at both will put us in a better position to state an alternative understanding of fruitful uses of thought experiments.

Let's look at Wilkes's analysis of Plato's famous thought experiment of the *Ring of Gyges*. In *Book II* of *The Republic*, Glaucon challenges Socrates to show that justice is good both in itself and in its consequences. According to the view that Glaucon is expressing, justice is done only because the consequences of being caught outweigh the profit from doing injustice. The *Ring of Gyges* thought experiment is supposed to demonstrate that no one would remain just if he possessed the ring that would allow him to evade punishment by becoming invisible. Glaucon thinks we would all agree that both the just man and the unjust man would act the same way, had they each been given the ring (Plato, 357e–360d). The idea is that justice is not good in itself, but only instrumentally, and this is expressed by the intuitions revealed by the case. According to Wilkes, here "the imaginary state of affairs is the invisibility; one conclusion *may* be that morality must be based ultimately on self-interest" (1988, 5).

Wilkes, predictably, thinks that this thought experiment is inconclusive because not all the relevant background is spelled out. Specifying the full background adequately presupposes answering at least some of the following relevant questions, she thinks: "If you are both invisible and intangible, could prison walls hold you? And if they could not, could *you* hold a gun, or a caseful of banknotes? Again, would others know that one owned such a thing[?]," etc. (Wilkes 1988, 11). As we ask these questions, we are supposed to see that the impossibility here is destructively relevant to specifying the

[22]Saying this does not exculpate Wilkes's opponents, of course, since both parties are assuming roughly the scientific model.

background by answering such questions. At some point in this questioning, as far as I understand Wilkes, we will see that there is a (possibly contradictory) tension between the imagined invisibility and what one can expect to accomplish while being invisible. I take it that Wilkes thinks that either one will not be able to enjoy any of the benefits of being invisible and intangible or one will not in fact avoid punishment. Either way, the thought experiment does not show that both just and unjust act out of prudential considerations only. So, we cannot draw conclusions from this flight of imagination.

As Cora Diamond notices, this analysis really seems to be missing the point of what the thought experiment may be doing *in Plato's discussion taken as a whole.* The exploration of further properties of the invisible criminal seems irrelevant, if we reflect on the context of the discussion of justice that spans the entire dialogue. Furthermore, Diamond thinks that a sophist just needs an idealized case in "which there is a firm and entirely justified belief in undiscoverability of the 'test person,' reflecting such belief" (2002, 232). The crucial point is that Wilkes ignores the fact that for Plato, the motivations and the behavior of the just person will only be appreciated by the end of *Book X* of *The Republic.* Given a proper understanding of justice, according to Plato, we will see that the consequentialist analysis that Glaucon suggests will not be applicable to the person who is just in Plato's sense. So, unlike Wilkes, Plato does not view the thought experiment as a problem (to which both the sophist and Socrates give different answers) to be resolved at the time of its appearance in the text, but rather a *tool for exploring* issues of justice for the duration of the work.

In her general discussion of thought experiments, Diamond (2002) uses Paul Humphreys's (1993) terminological contrast between "well-posed" and "exploratory" problems to provide a different assessment of the value of thought experiments. The main contrastive feature is this: well-posed problems presuppose a particular understanding of the method of the problem's resolution and a unique solution to the problem, while exploratory problems are open in this respect. Diamond, following Humphreys, thinks that at least some thought experiments have a dialectical function: they modify the original assumptions

that went into constructing the scenario, provide additional arguments to support this or that way of thinking about the setup of the problem, etc. Many new insights can of course come from reexamining the familiar "facts" assumed in the setup of the thought experiment (Diamond 2002, 235).[23]

This way of understanding thought experiments would be missing entirely if the sophist simply accepted Wilkes's criticism and retracted his argument that morality may be based entirely on self-interest. Having done that, on the other hand, he could certainly have other arguments to the same effect, still thinking that the problem of justice, as he presented it, is well-posed. However, that would not get to what is at issue for Plato, according to Diamond, namely the idea of a connection between a well-ordered soul, political justice, and the entire educative order that fosters such understanding. Given that the issues at play for Plato are so complex that they preclude taking it for granted that the sophist's question is to be accepted on its own terms, we should take this episode to illustrate that there may be disagreements about *what exactly the relevant issues are for the purpose of evaluating a given thought experiment*, disagreements about "the character of a problem" (Diamond 2002, 243).

What can we learn from these two different assessments of the case? Diamond's analysis challenges Wilkes's assumption that thought experiments need to be well-posed problems in order to succeed. She challenges the assumption that we may know in general, without looking at the particular circumstances of the use of each thought experiment, whether it is a well-posed or an exploration problem.

One may think that this opens up a general response to Wilkes from anybody who uses thought experiments in personal identity: to construe them as exploratory devices. And there may be some truth to this response. Thought experiments are typically occur within some larger argumentative framework, some of them are modified later on, and so on. Maybe much of this can be packed into saying that thought experiments are "defeasible."

[23] See Kuhn 1975 for an earlier discussion of this possibility.

This move is too quick, however. I think that a more accurate picture is that the line between using thought experiments as well-posed problems and exploration problems is often blurred, crossed at convenient points, ignored, and so on. For example, I maintain that Wilkes's criticism applies to the way Parfit approaches the issue. (At this point, I ask that the reader grant this for the sake of argument, and I discuss particular cases in later chapters.)

Still, don't we have a kind of response to Wilkes? Shouldn't we grant that thought experiments that are not assumed to be well-posed problems can escape between the horns of Wilkes's dilemma? Since they are not meant to be well-posed problems, they are not meant to fit the scientific model, and we can continue to explore the relevant issues with thought experiments. Now, there will be several options about moving forward armed with this proposal. One is to go through particular thought experiments that Wilkes criticizes to see whether they are exploration problems or not, and address various problems that will inevitably come up in that discussion. This may include some of the options I mentioned earlier, such as getting into the details of particular practical concerns and their relation(s) to metaphysical questions of identity, for example. It may turn out that Wilkes's analysis may be correct for some thought experiments and incorrect for others. I won't pursue exactly this line of thought in the thesis.

However, there is another related idea I wish to pursue, and it is connected to the issue that all parties may have to confront, which is expressed in the spirit, if not in the substance, of Wilkes's criticism. I am convinced that Wilkes's discussion is rather useful, despite the exploration response. Suppose that Wilkes's treatment of the *Ring of Gyges* is easy to criticize. The case is not in fact central to Wilkes's general line of criticism, I think. Maybe Wilkes meant this example as a rhetorical device, and its failure does not detract from the usefulness of her discussion. The more general point that Diamond *shares* with Wilkes is this: we cannot take for granted whether or not this or that kind of stipulation is going to be useful. And this general idea brings back the problem of the fantasticity of the thought experiments with which we started, and that problem has not

disappeared with the exploration response.

Let's focus on the particular domain of questions of personal identity. Suppose we are after the heart of our understanding of what we most fundamentally are—"the heart of our condition," as Wilkes puts it. The exploration response has put a foot in the door of a way to use thought experiments. But this is not enough. In addition, we need some indication that this kind of exploration has the cognitive value of the sort that is useful for personal identity. It is not an accident that Wilkes tied the success of thought experiments to a model from science, which fits very neatly into the category of well-posed problems.[24] The scientific model aims at understanding how, as Tamar Gendler has put it, what is not the case can give us knowledge of how things are (2002). Once we move to the exploration model, we should look for an alternative account of how it supplies knowledge or has cognitive value.

Diamond (2002) says that some thought experiments are neither fairy tales nor well-posed problems (243). Whatever one thinks of the exploration model, one has to confront Wilkes's point that there is something problematic about the fact that these may just become *stories*, and admittedly there is a problem about understanding the cognitive value of fictional stories. That is, even once we give up on thinking of all thought experiments as well-posed problems, we have a more specific problem at hand: thought experiments may be cognitively stimulating, but so are other things. As you recall, Wilkes's suggestion was to turn to actual puzzle cases because they present us with a full range of troubling situations, and we don't have to make up the background: it is the background of our world, and we know where to look for solutions. Adopting the notion of exploration, then, we have yet to show that there is something significant about stipulated, fantastic, impossible, and so on, scenarios. Unless we say something more specific about the cognitive value of fictional discourse (including thought experiments), our exploration may be unmotivated.[25]

[24]See John Norton's papers, arguing that scientific thought experiments are disguised deductive or inductive arguments. Norton 1996, 2004.

[25]In general, this is one of those cases, in which it is a pressing issue for philosophers to account for why their activities take this, rather than that, shape: why they use imagination rather than science, in

All this may be too quick. It may be enough for a defender of the exploration program to point to various successes of thought experiments in getting to some interesting results, as I said earlier. But I think we need something stronger than this—an account of the cognitive value of fictional discourse that gives us more than the pragmatic success of some past instances of thought experiments. If we can get to such an account, we may yet devise the possible escape between the horns of Wilkes's dilemma via exploration. This suggestion opens up the line of thought that may have been troubling us from the start. Why should we assume Wilkes's formulation of the dilemma, that either philosophical thought experiments are like science or they cannot give knowledge? If we can give an account of the cognitive value of fictional discourse—of which thought experiments are just one example—in ways other than those that make it mimic science, then we can avoid Wilkes's conclusion that thought experiments are not useful in personal identity.

Thus, one may also come to resist Wilkes's dismissal of the use of thought experiments from a different, and perhaps unexpected, angle. Wilkes's conclusion that thought experiments are not fruitful in philosophy is offered to us against the background of the long-standing cultural practice of hearing, speaking, and writing about doubles, disembodiments of all sorts, switching and exchanging bodies, and so on. We may dismiss this practice as having no significant bearing on our self-understanding, but a less dismissive approach will attempt to give these stories some constitutive role in forming our thinking about ourselves. Philosophy does not start in a cultural vacuum, nor is it our only source of knowledge outside of science. To dismiss these stories as useless for our philosophical understanding because they cannot be modeled after thought experiments in the physical science would be to reduce all learning to one model and all thought experiments to the same function. I suggest that a different understanding of the function of thought experiments or the question they are trying to answer will allow us, among other things, to give them an adequate place in our philosophical culture and provide the proper understanding of the importance of the background. In the next two chapters, I look at literary

this case, why this does not result in a more explicit recognition that this is a rather literary enterprise, and so on.

fictions and thought experiments in bioethics to explore a different way of understanding imaginative exercises as sources of cognitive value.

In the rest of this chapter, I suggest we go back to Wilkes's own discussion of literary examples to say a bit more on the issue of what assumptions about the function of thought experiments and philosophy in general Wilkes makes, and perhaps see if the seeds of a different understanding of thought experiments are already planted in her own discussion.

Recall Wilkes's remarks about the reasons why exploring what she calls logical possibility is not fruitful for philosophical purposes. According to her, this is so because logical possibility, as exhibited in literary fantastic stories, cannot on its own warrant any useful conclusions: either what we get is too general, or we cannot sort out "to which of the variable and varying factors to ascribe" the ceteris paribus conditions (Wilkes 1988, 45). In contrast, legitimate philosophical thought experiments, Wilkes tells us, aim at allowing us to tellwhat we would say "precisely *in* the world as we know it from our scientific theories—so that we can explore the ramifications of the concepts in question" (1988, 45).[26] Moreover, Wilkes thinks that literary fictions are written for *entertainment*. According to her, "the author of fantasy sets out for us a new framework within which the events taking place are (given a dollop of suspended belief) intelligible. He is not setting out to enable us to draw conclusions about our theories and our concepts"(Wilkes 1988, 45-46).

This much may be true: when we sit down with our fantasy or science fiction novel we are not—and maybe the author is not—explicitly aiming at working out some theories of personal identity by generating intuitions about answers to identification or reidentification questions. On the other hand, being entertaining does not preclude engaging with the deep concerns with and questions of identity in our own lives. The sharp division between what is philosophically fruitful and what is entertaining may just be the outcome of Wilkes's focus on the scientific model and her choice of examples. While she discusses

[26]Haagqvist agrees that the thought experiments he would allow aspire "to test some hypothesis or theory" (1996, 15).

Carroll and Tolkien, she does not mention Kafka or Mann.[27] But surely we do not want the force of our philosophical conclusions to rest only on the selected examples.

Wilkes is clearly aware of the way in which we can challenge the established understanding of this or that feature of our world by means other than those of searching for incompatible evidence or giving an argument. As I am suggesting, following Humphreys's and Diamond's idea of exploration,[28] we can challenge the terms of the evaluation, the very criteria by which evidence is admitted or dismissed as relevant, depending on the purpose of our investigation. (The disagreement, that is, may be about the character of the problem, before it can be about how to resolve it, since our understanding of the latter presupposes the former.) It is therefore surprising that Wilkes's own discussion of the notion of relevance proceeded, for example, without opening up to a more flexible view about the potential role of thought experiments.

I think this narrow vision of the possibilities for the fruitful use of thought experiments stems not only from Wilkes's responding to a set of texts whose authors explicitly or implicitly seem to invoke science as their model, but also from a particular view of the demarcation between scientific and literary-fictional discourses along the lines of their ability to speak about reality. The assumption is that the latter are deficient in this respect simply in virtue of being fictional, i.e., made-up. In the next chapter, I explore different approaches to the cognitive value of fiction. Ultimately, I will argue that we can benefit from understanding the cognitive value of thought experiments by first looking at the cognitive value of literary fiction, and this realignment of allegiances (drawing analogies to fiction rather than science) will require rethinking what we do with thought experiments and what we can learn from them.

[27] The latter have a reputation for being more "philosophical," as anecdotal evidence suggests to me. And, in fact, some would argue with her about both Carroll and Tolkien.

[28] Also see Bokulich 2001, and Kuhn 1975.

2.6 Conclusion

I devoted the bulk of this chapter to Kathleen Wilkes's important criticism of thought experiments in personal identity. Wilkes tied the conditions of success for thought experiments in personal identity to their mimicking scientific cases. I then suggested that there is something problematic about Wilkes's assumption that if thought experiments in personal identity do not fit the scientific model, then they are merely stories and cannot be cognitively significant for our understanding metaphysics of persons. Granting Wilkes's conclusions about the adequacy of the scientific model as applied to philosophy, we can start by considering Diamond's suggestion that at least some thought experiments may play a different, exploratory role. If some thought experiments are meant to be exploratory, it seems that we can avoid Wilkes's complaints because exploration may not require such a stringent understanding of the relevant background as Wilkes demands. Commenting on this move, I suggested that Wilkes's work contains another assumption that serves as the point of departure for the way for the rest of the thesis: that fictional stories cannot be cognitively significant for metaphysics. The assumption is that if we cannot bring thought experiments under the rubric of science, they are *merely* stories, and it is even worse in the fantastic cases, since in those imagination supposedly runs free. To respond to this objection, we need an account of the cognitive value of fiction that is not like the accounts given by science. This is the subject of the next chapter.

Chapter 3

The Cognitive Value of Fiction

3.1 Introduction

I ended Chapter Two with a discussion of Wilkes's views about literary fictions. According to Wilkes, philosophical thought experiments that do not fit the scientific model begin to resemble fictional stories, compromising the distinction between serious philosophical engagement and philosophers' "play." If we are going to defend the use of such stories in philosophy as a methodological tool, we need an account of the cognitive value of fiction.

Wilkes's sharp demarcation between knowledge and fiction is understandable. After all, fiction does not seem to have the markers we typically associate with the production of knowledge, such as evidence and hypothesis generation, theoretical construction, explicit argumentative strategies and structures, and so on.[1] This view can be expressed like this: fictional discourse is not constrained by the same rules we associate with theoretical investigation, scientific or philosophical. Think of how often we hear that fiction is associated with *escaping* rather than investigating this world: intuitively, worlds made of words have a very different status than those we can touch and experiment with. Falling in love with a literary character is different from falling in love with an actual person; a death of a character in a book has rather different consequences from an actual death. Assuming that philosophers are in the business of truth-seeking or knowledge-seeking,

[1] These remarks may have obvious counterexamples in various works of art that play with adopting different styles of non-fictional discourse for the purposes of literary production. I bypass this matter entirely.

broadly understood, turning to fiction for knowledge sounds somewhat paradoxical.

At the same time, Wilkes's remark on the epistemological value of literary works misses entirely what John Gibson calls the *humanist intuition*: the idea that despite being fictional, literary works are also *our-worldly* in the very direct and immediate sense. We are convinced that one of our culture's ways of learning and acquiring knowledge is through producing and reading fiction (Gibson 2007; also see Harrison 1991).[2] Some philosophers think we can learn quite a lot from fiction.[3]

In this chapter, I will be occupied with the apparent tension between knowledge and fiction and the question of the cognitive value of fiction. I restrict my remarks to narrative fiction, broadly construed. Moreover, my more narrow interests lie in understanding what insights the more fantastic variety of literary fictions can offer us. Stories of incredible transformations and fantastic possibilities are wide-spread: adventures of doubles, accounts of invasion or possession by spirits, discussions of body-swaps, head transpositions, or humans acquiring non-human form, are all familiar enough. Later in the chapter, I discuss some examples of what I call the 'literary counterparts'[4] of philosophical thought experiments: Franz Kafka's *Metamorphosis*, Thomas Mann's *Transposed Heads*, Stanislaw Lem's *Solaris*, and Karel Capek's *The War with the Newts*. Even though the sample is rather small, I hope it will make my point. While these stories are of interest to anthropologists and psychologists, who study the cultural significance of symbols, mechanisms of knowledge transmission, and other questions, these stories, arguably, are directly connected to perennial philosophical themes such as the relation between mind and body or spirit and flesh, the role of memory as opposed to the role of current circumstances (the role of the past in shaping the future), uniqueness as opposed to similarity, and many

[2]There is the weight of the tradition that has defended the idea that we can learn from literature: from Aristotle, through Horace, and Dr. Johnson, through realist novelists, to Marxist readings, and so on (Lamarque 2006, 131).

[3]Some philosophers think we can at least learn the following: factual information about the actual world contained in fictions; understanding of general principles about morality or psychology, explicitly stated or implicit in the works; categorical understanding that gives us novel ways of categorizing the world, including acquiring different skills and strategies for understanding the world; and affective knowledge given by the ability of literary fictions to put us in another's shoes (Davies 2007, 145–146; Novitz 2008, 345–346,349).

[4]Thanks to Marya Schechtman for this useful term.

others. I hope that examining the ways in which literary fiction can give us knowledge of this world can illuminate important ways of understanding the cognitive value of thought experiments in personal identity.

The relevance of looking at fiction to our overarching question will not be given in full details until Chapter Five, but I should say something about the function of this and the next chapter in the overall structure of my argument to forestall a possible misunderstanding. My proposal is methodological. I am not arguing that reading any of the stories I mention will address the questions in which philosophers are interested any better than will philosophical discourse. The proposal is not to replace philosophical thought experiments with literary fictions or to ask philosophers to become better writers. Instead, our interest lies in understanding the resources that make the fictional details of these stories cognitively significant. Because fiction's ability to present us with fictional worlds can be illuminating in different ways than scientific and philosophical investigation (at least as it is traditionally understood),[5] if we understand how literary fictions can be cognitively significant, we can use these insights to look for similar resources in philosophy. I do not suggest that we replace the content of the philosophical thought experiments with the content of their literary counterparts. Instead, I am interested in the ways in which the resources of fiction-making can be illuminating beyond the ways in which the intuitions generated by philosophical thought experiments are used in personal identity.

Of course, nobody should deny that literary works are cognitively stimulating, or, to put it bluntly, that literature makes us think. Following Eileen John (1998), we can distinguish between stronger and weaker theses of the cognitive significance of literature. Thus, the weaker reading is in terms of stimulation: x is cognitively significant as long as x makes us think about the world in important new ways. This is too vague to provide anything like a model of how specifically literary elements as part of the fictional narrative are knowledge-capable. What we hope for is something stronger: an account of the cognitive significance of literary fictions that locates the sources of knowledge in

[5]Some philosophers may disagree, but then this may be a sign of their being convinced of what I am saying here.

the literary elements themselves (John 1998; Gibson 2007, 2009). If we do not aim for the stronger claim, we lose the reason to think that there is something special about literary counterparts of philosophical thought experiments beyond their amenability to philosophical treatment.

I proceed as follows. In section 2, I state three standard arguments against the idea that literary fictions can be cognitively significant. In section 3, I turn to one strategy of defending fiction against these arguments, namely the idea that the role of the literary fiction is to present some raw material for further philosophical "translation." While these efforts are driven by the desire to expand the philosophical repertoire, these views may threaten to make literary fictions themselves disposable. Since I think the insight of the humanist intuition is to locate the cognitive value in the fictional elements such as they are, in section 4, I turn my attention to the views that seek the contribution of the literary works to our knowledge in the process of engaging with the work as fictional, and not some backhanded way of doing 'proper' philosophy. I follow Martha Nussbaum, Cora Diamond, and John Gibson, among others—however (un)comfortable each may feel in each other's company—to sketch the view that close attention to the fictional details of literary works brings to the surface the sense and significance of the fundamental practices of our culture. To conclude, I add a discussion of some specific literary fictions to substantiate my points.

3.2 Against the Cognitive Value of Fiction

Recall the paradoxical sound of the idea that fiction can give knowledge. The following arguments challenge the idea that we can connect the two. The first two deny that the structure of fictional discourse is in the same playing field or in the same weight category as the structure of truth-bearing discourses; the third argument accepts the idea that we can get knowledge from fiction, but questions its significance.

3.2.1 The No-Evidence Argument

The no-evidence argument attacks the idea that fictional discourse is in the same business as theoretical-evidential discourse. Fictions do not advance hypotheses, collect or present

evidence to support these hypotheses, or aim at systematizing our knowledge of the world, etc. The standard of evaluation of the success of a literary work is not based on accuracy or successful mimicking of the disciplines that are investigatively commenting on this world—the sciences, for example. Literary works have a different relation to the world than that of truth, reference, implication, etc. (Gibson 2007, 2, 29, 30, 34; Carroll 2002)

A closely related idea deserves special mention here. Various proposals about how to understand the "truth" in "truth in literature" have been met with serious objections. For example, one of the early proposals—the so-called "propositional theory of literary truth"—can be stated like this: "the literary work contains or implies general thematic statements about the world which the reader as part of an appreciation of the world has to assess as true or false" (Lamarque and Olsen 1994, 325; Kivy 1997, 121). But the idea of trying to understand literary works as straightforwardly *implying* truths about the actual world (on some suitable understanding of 'implication' as material conditional, inductive inference, or counterfactual dependence) is something of a non-starter. As Sirridge argues, if we relax the notion of implication/containment here to mean something closer to 'elicit,' 'show us,' 'suggest,' 'can be symbolically paraphrased as' and so on, the main problem should be apparent if one realizes how context-dependent such relaxed notions are. We simply do not have some general way of distilling the drop of truth from what is given to us in the literary potion. Needless to say, there will be disagreements on what these truths are, and some of what can so be distilled will in fact be false. (Sirridge 1975). Notice that moving to a view that the truth that literature can provide can be explained as subjective truth from the standpoint of the narrator does not help: it fails to give reasons to take these truths as generalizable. It is similarly not enough to say that their value comes from the fact that we can occupy different perspectives while confronting the work of art: increasing the number of wrong takes on the matter presumably won't make it right, and we need some independent reason to know that the views here are expressing genuine truths (see Lamarque and Olsen 1994).

Even if *some* truths can be said to be implied by literary works, where this is un-

derstood in some suitably weakened way, their evidential support does not come from the fact that they have been put on paper by this or that author—their warrant is not secured by their being works of fiction (Carroll 2002, 5). Moreover, the question of quantifying those implied truths is problematic: if we are talking about genuine generalities, we should have controlled studies, peer reviews, and so on (Stolnitz, 339). And if anyone were to test those generalities, again, the evidential support would not come from the fact of fictionality of the work. Moreover, there seems to be no special expertise involved in learning how to acquire this kind of knowledge (Stolnitz, 341). Furthermore, what truths those are to be confirmed is by far not clear.

There are things to say about each of these points, no doubt, but since my purpose is not to address these arguments, I hope that merely listing the problems already shows just how serious the challenge is.

3.2.2 The No-Argument Argument

As the previous argument may have indicated already, if fiction is not in the business of advancing and evidentially supporting knowledge-claims, then it cannot be the kind of activity that is in the business of generating knowledge via standard techniques of knowledge production: argument, analysis, debate (Carroll 2002, 6; Gibson 2007, 2,30). Furthermore, then, its standards of success cannot be tied to providing arguments, generating truths, and so on. Again, the idea that fiction is truth-generating is suspicious (Carroll 2002, 6; Kivy 1997).

3.2.3 The Banality Argument

Very roughly, even if literary works can be thought of as implying some truths about how things are in the actual world, these truths are trivial. (This should be apparent for some of the reasons stated earlier: the scope of generalization is unclear; the experts are nowhere to be seen [Carroll 2002, 4; Stolnitz ,337; Gibson 2007, 10].) The claim is that what art does is merely a reappropriation or reusing of truths that are already available to us. It would be strange to think that we do not know the pains of betrayal or the

beauty of friendship prior to reading Shakespeare or Dostoevsky. Now, it may be that art presents these truths differently, puts them in different contexts, or does some other such thing. However, this does not show that the source of knowledge is to be found in the fiction itself.

3.3 'Instrumentalism'

The common theme that runs through all three of the arguments presented above is the assumption that to the extent that the fictional discourse does not bear the marks of theoretical (scientific/argumentative) discourse, it is deficient in terms of its ability to generate knowledge. In this section, I look at two related approaches to understanding learning from fiction, which attempt to address these arguments in the 'instrumental' fashion. In very broad strokes, literary fictions can be interpreted as containing arguments, making claims or hypotheses about the world, and so on. However ingenious, these attempts suffer from a serious problem.

Noel Carroll makes the following argument to the conclusion that fiction is cognitively significant. Philosophers use thought experiments; thought experiments can withstand the three arguments above and do produce knowledge; if we show that literary works are similar to thought experiments, they, too, can be similarly cognitively significant (Carroll 2002, 7). Carroll takes for granted that thought experiments are widely used, and he does not seem to be bothered by the problems I have been discussing in Chapter Three. This in itself is not a problem for Carroll since all he needs to show is that *if* philosophers are happy with thought experiments, they can be happy with fictions in general. We can approach Carroll's argumentative strategy from two directions. First, we should question the usefulness of the philosophical thought experiments along the lines familiar from Chapter Two. Second, we should wonder whether thinking of literary works as thought experiments helps the philosopher at the expense of the work itself. Let's discuss these points one at a time.

Carroll's conception of the thought experiment is fairly standard: it is a tool of con-

ceptual clarification, using which we can trace implications, refute various modal and other claims, provide counterexamples to overgeneralizations, explore our commitments and so on. (Notice that the generality would include both the hypothetical and the fantastic cases.) In particular, thought experiments: 1) excavate conceptual refinements and relationships; 2) bring to the surface implicit conceptual knowledge, and sometimes shift our conceptual schemes (Gendler 2000, 2, 25; Kuhn 1975); 3) in general are used "for the purpose of framing probing, and/or challenging definitions, for testing ways of setting up a question or a problem, for making precise distinctions revealing adequacy conditions, tracing entitlements and inference patterns, proposing possibilities proofs and assessing claims of conceptual necessity" (Carroll 2002, 8). According to Carroll, it does not matter whether the thought-experimental examples are fictional or actual since we are here working with concepts: we get the "mind moving over its conceptual map" (2002, 8).

In short, thought experiments explore our conceptual commitments by presenting us with fictional scenarios and result in conceptual reorganization. They are presented in incomplete form that the readers fill in with their antecedent knowledge. In complete form, then, these conceptual structures function like arguments, and so they are immune to both the no-evidence and no-argument arguments. As for the banality argument, Carroll thinks that it depends on the depth of conceptual transformation in question: it may be profound, of course, depending on the audience.[6]

I have reservations about Carroll's argument. First, if Carroll's strategy is to appeal to the general acceptance of the methodology of thought experiments among the members of the philosophical community, we can simply point out that agreement on the cognitive value of thought experiments is neither universal nor without problems. As I argued in Chapter Two, Wilkes's criticisms challenge the coherence of at least the fantastic variety of thought experiments. Even though some of the thought experiments that Carroll has in mind may clarify our concepts, the fantastic ones are harder to tame. Simply on these grounds, Carroll's analogy can be restricted. How reliable can our intuition be—even if

[6]Similar remarks on the relevance of what audience you are presupposing are found in Kivy 1997, 127-128.

it is about conceptual refinements—when we are asked to think of the world in which we split, for example? Getting the mind moving is no easy matter, but stopping it from idly spinning is no easier task. The fact that we are exploring conceptual commitments does not mean that there are no constraints on the intelligibility of the enterprise: in many cases we should not be so sure about the relevance or significance of probing our conceptual scheme with far-fetched scenarios.

But even though we may accept the idea that some thought experiments can be cognitively significant as conceptual refinements and clarification tools, we may still question the *analogy* that Carroll draws between philosophical thought experiments and literary works. More precisely, Carroll claims that, under a given interpretation, literary works can be seen as making claims about possibilities, making clarifications, distinctions, and so on (2002, 11). Some of them "cultivate our grasp of what is known with finer distinctions, while others may possess comparable structure to various thought-experimental techniques of variation: they present within themselves various takes on the situation, and these structures get picked up by their audience" (Carroll 2002, 11). In sum, "literary fictions then can afford knowledge of concepts, such as concepts of virtue, by stimulating the reader to an awareness, through reflective self-analysis, of the conditions, rules, and criteria for the application of said concepts" (Carroll 2002, 14).

While what Carroll says sounds unobjectionable, one may think that this description of learning from literary works is overly philosophical: it in fact describes what the philosophers do with the literary work when they try to *use* it to their own advantage or in support of their theses. There may be nothing wrong with that, considered on its own, but one may also notice that something gets lost in the transition from a literary work to a set of cognitive structures present in the thought experiment. To put this bluntly, literature is not an argument. This returns us to the set of arguments I gave at the very beginning: they can be thought of not only individually as arguments against the view that literature can provide knowledge, but also as expressions of the idea that there is something really distinct (and valuable) about reading literary works as such, and that

we may lose sight of that aspect if we start doing philosophy with them, *to* them.

Carroll is aware of this objection, but points out that this should come as no surprise: given how integrated fictions are into our social life, the step between discussing the work itself and, for example, the virtues and vices at work in the literary work, is very short (2002, 16). In fact, the step from concrete immersion in literary fiction to a generalized discussion of moral virtues by a philosopher-critic is contained in our practices of reading, writing and evaluating literary fictions. This is how we teach children morality; this is what we do in literature (and of course philosophy) courses, etc. According to Carroll, even though the poet may speak of what the Muse whispers, what we care to hear often depends on what we can do with what is said: in the classroom, or in the kitchen. The interactions that constitute the institution of literature—between the work, the audience, the critic, the philosopher, and so on—afford many examples to substantiate Carroll's starting analogy. In practice, we can take literary works to contain the same kind of structures as philosophical works, and argue for that.[7]

Having said all this, Carroll recognizes that there is a big difference between the philosophical thought experiments that tend to be abstract and schematic and the literary examples that are full of details and are much more context-driven. According to him, the richness of detail in literary works—now pictured as implicit thought experiments—is required by the complexities of thought-experimental problems. Discussion of the subtleties of virtues and vices, for example, depends on all sorts of hidden nuances, and much more elaborate detail is required to discuss them than is standardly given in philosophical abstractions (Carroll 2002, 19). Philosophers usually say that more detail can be supplied upon request, and we saw that Carroll himself thinks that the details are implicit in the abstract descriptions. The concrete details of literary presentation, according to this reading, are "grounds for appreciating them as thought experiments that have *special cognitive requirements and advantages*" (Carroll 2002, 19, my emphasis). This is a telling

[7]Compare Lamarque and Olson, for example, who distinguish between philosophy *in* literature and philosophy *through* literature; the latter seems to signify the same idea of a fully worked-out philosophical theme exposed in literature (1994, 391). But also see Danto 1984.

remark, and we will return to it in a moment.

A similarly cognitivist approach in defense of a weak version of the propositional theory of literary truth is advocated by Peter Kivy. He writes: "Some fictional works contain or imply general thematic statements about the world that the reader, as part of an appreciation of the world, has to assess as true or false" (Kivy 2007, 122). According to Kivy, the problem with our starting skeptical arguments is a limited understanding of what literary experience comes to. He argues that works of fiction have a certain (reflective) afterlife, during which the reader can confirm or disconfirm the hypotheses she may find "live" or "dead" to her, while she is reading or taking a break from, or thinking about, etc., the work. These reflections, according to Kivy, cannot but in some sense be connected to thinking of the hypotheses along the lines of truth or falsity (2007, 126, 132, 135).[8]

The ideas expressed in Carroll's and Kivy's views seem to assume the primacy of the methods of philosophical analysis. But this approach does not cover all possible ways of thinking about the connection between literary fictions and thought experiments; moreover, it may foreclose other possible ways of thinking about the issue. Here is how John Gibson describes one problem of this kind of making philosophy out of literary fictions. Recall from the introduction to this chapter his way of putting the humanist intuition: we are trying to reconcile the fictionality of literary works with their our-worldliness, i.e., we are trying to do justice to literary fictions as such without either making them into some pure creation without significant connection to the truths of this world, but also without making fiction into a mirror of this world. Carroll's and Kivy's approach is, roughly, what Gibson calls "indirect humanism": "it brings the literary to bear on the our-worldly by exploring our ability to apply aspects of the content of a literary work to extra-textual reality" (2007, 18). The indirect humanist is talking about using the literary instrumentally in order to acquire knowledge about the world. But this does not tell us about other ways in which the specifically fictional elements of the

[8]The terminology of 'dead' and 'live' hypotheses is borrowed from William James's "The Will to Believe."

work can open up the view of the real. If we always look to the fictional to rework it into something other than fiction that will connect the work to the world (an implicit argument for a hypothesis, for example), we have not explained the cognitive significance of the fictional as such (Gibson 2007, 35); i.e., we have not spelled out what engaging with specifically fictional elements of the literary contributes to our knowledge of the world. Instead, we may have confirmed that literature makes us think, and sometimes the ways in which it makes us think can be presented as (implicit) thought experiments.

In sum, the "instrumentalist" proposals above emphasize making fictions into suitable targets of theoretical talk about conceptual commitments, implications, and other philosophical machinery. But why do we think that the fictional elements can do that kind of thing? Why are we drawn specifically to fiction, including fantastic and experimental kinds, to probe our conceptual commitments? What is it about the fictional specifically that can even suggest that this kind of looking may be fruitful?

As we saw, Carroll's closest suggestion to appreciating specifically literary elements of fiction comes in the form of acknowledging their concreteness and detailed character. Carroll's take on it, as we saw, is reminiscent of Wilkes's thinking that additional details of literary works supply the contextual background to make the (implicit) thought experiments more precise. Carroll (2001) calls his general view "transactional" or, as he prefers to call it, "clarificationism":

> [c]larificationism does not claim that, in the standard case, we acquire interesting, new propositional knowledge from artworks, but rather that the artworks in question can deepen our moral understanding by, among other things, encouraging us to apply our moral knowledge and emotions to specific cases. For in being prompted to apply and engage our antecedent moral powers, we may come to augment them. (283)

Literary works, then, provide important focis for the exercise of our abilities and the clarification of our concepts. We train and refine them in the process of engaging with literary fictions.

I agree that if we are going to find something insightful about fiction's contribution to our knowledge of this world, we should pay attention to the role that concrete fictional

details play in this understanding. But as I have been suggesting, we may do a disservice to literary fictions themselves by *exclusively* (mis)representing them as something else: arguments, thought experiments, theoretical aids, tools for clarification, and so on. Of course, I do not deny the value of such approaches, but I do want to question whether they exhaust the possibilities of our understanding how a work of fiction can have cognitive value. In what follows, I consider several proposals about understanding the cognitive value of fiction without assimilating literary works to thought experiments or arguments. This discussion will provide conceptual space and vocabulary for talking about openness with respect to our philosophical methodology.

3.4 The Cognitive Value without Assimilation

Behind the idea that to understand the cognitive value of fiction we need to assimilate it to philosophy may be a limited conception of what can count as legitimate ways of knowledge production. One of the reasons that some serious philosophers of literature warn us against the moves to assimilate literature to philosophy along the lines of the previous section is that these ways misrepresent literature's own special mode of knowledge production (e.g., Lamarque and Olsen 1994). In this section, I look at several views that in various ways seem to be opposed to the sorts of assimilation of literature to philosophy in ways we just encountered.

In several papers, Martha Nussbaum argues for the particular place that literary works must occupy in a more complete understanding of ethical reflection. Her picture of ethical activity involves and in some sense prioritizes emotional engagement with particular characters, which finds its natural home in works that are usually considered literary in form, over the application of general rules (Nussbaum 1990, ix). Let's spend a little time with the details of her account.

Nussbaum asks: If our question is to understand what it is to lead a good life, and ethical reflection discloses some ways of answering this question, what should be the source of our reflection? In particular, why should we think that literary examples are better

suited for this type of activity? Nussbaum says that "[s]chematic philosophers' examples almost always lack the particularity, the emotive appeal, the absorbing plottedness, the variety and indeterminacy, of good fiction" (1990, 46). Since novels are typically open-ended, highly detailed and complex, and since they display both the inconclusiveness of live choices and the significant consequences of any choice, they have a closer affinity with our life-situation, in which genuine ethical choices are never like the cooked-up philosophical examples.[9] Applying these ideas to Henry James's *The Golden Bowl*, for example, Nussbaum argues that "flat-footed" philosophy, with its arguments and reasons, will always miss out on the "mystery, conflict, and riskiness of the lived deliberative situation" that show the value of good life-choices (1990, 142).

Novels trigger our emotional responses and put us in the shoes of, or along with, its characters, making the circumstances alive to us. By doing this, they represent the actual domain of ethical reflection more accurately. Nussbaum thinks that these works of art are indispensable to reveal that side of ethical deliberation (1990, 49). The picture emerging from these remarks is that of literature allied with philosophy, the philosopher providing both the explicit explanatory commentary and analysis of the passages in the novels and attending to the literary form in which they are presented. She says, for example, "To make room for love stories, philosophy must be more literary, more closely allied to stories, and more respectful of mystery and open-endedness than it frequently is" (Nussbaum 1990, 284). She quotes James's suggestion that literature is "confessedly tentative and suggestive rather than dogmatic," and appeals to the ideal of "attentive conversation" between philosophy and literature as a guide to this kind of discourse.

Nussbaum's efforts to extend the philosopher's attention to the particulars of literary presentation as more revealing than the schematic examples echo Carroll's comments. Thus, one might think that the difference between her project and mainstream philosophy is not so pronounced, and her view can be restated in the following way: introduce more

[9]Consider popular "trolley problems" and other typical hypothetical cases from ethics. Of course, the subsequent discussions of these caricature examples are always a bit more detailed, but that just goes to show that in this domain there are demands of more realistic presentation and attention to the context that may not be required and may in fact be distracting in other fields.

details into your philosophical scenarios, and then provide analysis of them, along the lines that are familiar to philosophers. This would be very much in the spirit of Wilkes, Carroll, and Kivy. (Or one may say that Nussbaum's remarks are at best suggestive, perhaps reflecting the set of values that she finds significant.)

This reading makes Nussbaum more traditional than she may be, however. There is also another side to her remarks. For example, she stresses the idea that

> one must be careful that the form and the stylistic claims of the commentary develop and do not undercut the claims of the literary text. And one must not rule out the possibility that the literary text may contain some elements that lead the reader outside of the dialectical question altogether; that, indeed, might be one of its most significant contributions. (Nussbaum 1990, 49)

If I understand this correctly, the dialectical dialogue between philosophy and literature should not proceed by simply adding literature to philosophy as giving us better (more detailed and concrete) examples with which to probe our conceptual commitments and entitlements in the familiar ways. I take it that Nussbaum urges us to consider whether seriously attending to the literary work will *change* our views of the goals, needs, and methods of philosophy.

But what exactly does this mean for a working philosopher? How is she to respond to this invocation to bridge the gap between literary presentation and philosophical methods? Going deeper into the details of Nussbaum scholarship is beyond the scope of this thesis. But her remarks are provocative enough and can be used to build up an analogue for a kind of attentive looking at what exactly happens in literature that utilizes the fantastic details without imposing a particular view of what a philosophical use of literature should be. Let me say more about this theme.

Closely related issues are raised by several exchanges between Cora Diamond and Onora O'Neill, prompted by O'Neill's review of Stephen Clark's book *The Moral Status of Animals* (Clark 1977, O'Neill 1980, Diamond 1991).[10] O'Neill's problem with Clark's view on extending moral status to animals is his failure to provide any metaphysical

[10] On subsequent developments of this issues see Mulhall 2009.

grounding for the extension of the treatment we reserve for humans to non-human animals. In his book, Clarke insists that such search for metaphysical grounding is a "paranoid fantasy" if it plainly denies our regularly observed emotional responses to animals. O'Neill thinks that in order for anybody to share Clarke's view on animals, these persons have to antecedently share Clarke's views. If they do not, they can only be convinced by rational argumentation, which alone overcomes the particularities of one's sentiments and can give us the desired universality (Diamond, 296; O'Neill 1980; also see Singer 1975). The danger of Clarke's appeal to sentimentality is its contingent and biased nature.

According to Diamond (1991), O'Neill's fairly traditional understanding of how we are to proceed in ethics—start with metaphysical features of reality in order to ground ethical views to avoid the supposed threat of ethical relativism—betrays an impoverished understanding of both philosophy and intelligence. Most crucially for Diamond, the assumption that we can avoid the threat of relativism only if we find some metaphysical features of reality on which to rest our ethics only precludes us from appreciating other resources we have against the purported threat of relativism.

We can gather from Diamond's remarks the most significant feature of the resources that Clark emphasizes, namely the attentive response to particular details of the world with the full range of one's intelligence and imaginative engagement. As Diamond (1991) describes Clarke's view,

> the transition [to such an attentive response] depends on our coming to attend to the world and what is in it, in a way that will involve the exercise of all our our faculties; and that religion, poetry, and science, if uncontaminated by self-indulgent fantasy, are the most important modes of thought leading to that kind of attentive imaginative response to the world. (296)[11]

Like Nussbaum earlier, Diamond does not deny that argumentation plays an important role in our seeing things right; nobody could deny that. But she denies the a priori assertion that philosophical procedures cannot appeal to *other ways of clarification and elucidation* if they are to count as genuine philosophy. Notice again that Carroll's view,

[11] The influence of Iris Murdoch's is significant in this line of thought [Murdoch 1971].

which we considered earlier, sounds rather similar. He wants to give an account of the cognitive significance of literary works by pointing out that philosophers have used techniques that are closer to what is going on in literature: narration and fictional imagination. There is an important difference, however, that we should not overlook: Carroll assumes that our aims and goals in this kind of activity are still probing and testing conceptual commitments, clarifying implicitly held views, and so on—i.e., that literary works can be seen as doing the set-up work for philosophy. In contrast, both Nussbaum and Diamond are trying to articulate a position of openness and exploration with respect to precisely the assumption that the works themselves somehow require the philosophical crutch in order to be *philosophically* useful.

Diamond draws our attention to the resources that are left out of the discussion by a particular picture of a philosophical work as an argumentative discourse that appeals to reason and that aims at grounding our convictions in some shared reason or fact. Following Henry James, she calls the resources that may be left out of the discussion by exclusive focus on argumentative reasoning, "habits of awareness, reflection, and discrimination" (Diamond 1991, 301). According to Diamond, these habits, especially the work that is done to develop them and cultivate them to their refinement, indicate and form the basis of the critical resources that are involved in how we attend to the world. As I understand her, these habits of discrimination, engagement, responsiveness, and so on, are what is given to us in the full description of what it is that a work of philosophy or literature *can* do in order to convince us. That is, we have options for assessment of the 'quality of thought' or 'quality of intellect' that do not have to reduce to assessment of argumentation and the availability of the metaphysical grounding for the philosophical view that is being defended.

I will quote her in full here:

> I have spoken of literary and non-literary works which invite us to respond emotionally or to take up some moral attitude or view of life; what I need to add is that such works may include in the 'invitation' an invitation to just the kind of awareness and critical reflection I have described. We are familiar

> enough with the kind of critical attention invited by philosophical argument, the kind of work demanded by it of the reader; but critical attention to the character and quality of thought in a work may be asked of a reader in many other ways as well. Further, a work may invite the reader to elaborate and develop a way of looking at things and to respond critically to it then as a possibility, perhaps leaving open in various ways how it is to be elaborated, perhaps incorporating any number of suggestions. (Diamond 1991, 303)

What is important in this passage are the notions of *invitation to critical attention* to the fictional work and *openness* to the possibilities that this attention can unearth. This seems to me rather different from (and richer than) the ideas explored in Carroll's view about the clarification of the conceptual knowledge we already have.

Perhaps I am quibbling with terms here, and perhaps we can see all of the options discussed so far—including Carroll's instrumentalism—on a continuum of modes of attending to the fictional elements in literary works. However, following John Gibson's work, what I think is distinct in this approach is the idea that the concrete details of fictional presentation that guide our immersion in fictional works, through the act of reading, present us with the *living world* of fiction, showing us how particular lives unfold. As Gibson puts the matter, the world that literary fictions present to us is a living world, and not a conceptual object, and "[w]hat literary narratives are able to do especially well is take the concepts we bring to our reading of a work and present them back to us as concrete forms of human engagement." This work of concretization of our knowledge shows us the world we understand as having value and significance in being an object of human care (Gibson 2007, 115–116).

The contrast of interest for us here is between engaging with literary fiction to clarify our conceptual commitments and engaging with it as a living world. The latter emphasizes the work that the fictional world does by exhibiting the *axiological dimension* of our understanding (Gibson 2007, 2009). It may be that the difference from Carroll's view is rather slight at this point, but it is a difference worth emphasizing. It lies in attitudes to the world that is being exhibited and developed before us in literary fiction. Diamond and Gibson think that we are dealing here with a living world to which we can be invited

and to which we can be open, a world that does not simply help us discover what we already know and make that knowledge more precise.

It may be tempting to think of the difference in terms of richness of detail along the following lines: philosophical thought experiments are abstract, and the role of incorporating literary fictions into the philosophical tradition is to make our understanding more nuanced. Carroll, and even Nussbaum, can be interpreted along these lines. However, richness of detail is not the main issue. As Gibson (2007) explains, the problem is that the instrumentalist's approach.

> requires that we sever the stances, themes, and perspectives we find in a work from their literary context and treat them as free-floating propositions, asking what they might tell us about reality if we disregard their place and function in the text and instead treat them as isolated assertions about the way the world is. (96)[12]

But what is lost through the philosophical instrumentalization? To explain the region of cognitive value that literary works can offer, Gibson utilizes Stanley Cavell's (1969, 1979) distinction between knowledge and acknowledgment. The idea of acknowledgment goes beyond that of knowledge in the following sense: one can be said to have knowledge that another person is hurt, hungry, desperate, and so on—i.e., one can be said to recognize the circumstances in which such terms apply (one can be said to have criterial knowledge)—without recognizing what kinds of action one is called on to perform when one knows such things (Cavell 1979; Gibson 2007, 2009). As Cavell (1969) put it in his discussion of skepticism, one can fail to recognize that a piece of knowledge 'makes a claim' on me to do something about it: to help, relieve the pain, and so on. Acknowledgment that something is to be done in the face of another's pain can be said to implicate and expose one in personal ways that go beyond the criterial knowledge of the concepts. Failure to acknowledge another's pain—in various ways—shows who one is in ways that

[12]This move is what prompts the skeptical arguments we started with. Once you have extracted the "propositions" of literary works from their embeddedness in the fiction, it is not surprising to discover how ill-suited they are to be part of something like a description of the world.

one's knowledge does not, it shows one's orientation to the world and others; it puts one in the space of values and moral understanding—the axiological dimension of our lives.

Based on this discussion, we can separate the criterial and the axiological understanding (Gibson 2007, 103; 2009). According to Gibson (2007), the concrete character of the living world of fiction places one in a position to appreciate what is at stake in how one's criterial knowledge is enacted and embodied in one's life, what it shows about the person in the world who fills the gap between merely knowing how to apply the term 'pain' and acknowledging the claims that another person's pain has on me: it shows how we choose to live with ourselves and others.

This distinction, then, can help us appreciate the difference in emphasizing conceptual knowledge as opposed to the vision of the world from an axiological perspective that a literary fiction may open up, and thus to clarify the ever-vanishing difference between Carroll's clarificationism and the axiological understanding discussed by Gibson and Diamond. According to Gibson, acknowledging is giving life to or exhibiting in a dramatic gesture the meaning and significance of our concepts as a living disposition embodied in our responses and actions to to others (2007, 108). Thus, acknowledgement completes our knowledge in the sense that it gives expression to the need for our cognizing the world to be embodied in our orientation to the world of others, the world of value (Gibson 2007, 111).[13]

Literary fictions, according to this line of thought, can be seen as inviting us to see how our knowledge can be fulfilled in this sense: it shows how literary work embodies what we know by giving more or less life to our knowledge (Gibson 2007, 111-112).

[13]One may think that the distinction between acknowledgement and knowledge presents two different *modes of cognition*. It is as if one first can have knowledge of the criterial kind, and then one can go on to apply it to living situations based on one's kind or cruel dispositions (i.e., one can move from knowledge to acknowledgment). But as Gibson (2007), following Cavell, points out, this is mistaken (111). Imagine a person who can expertly identify the signs of extreme pain in another person, see how she can relieve the pain, but is not at all stirred by it to do anything. Don't we want to say that there is something deficient in this person's response? Don't we want to know that the person does not *really* know that the other person is in pain? That is, criterial knowledge is not full knowledge without hooking up to the kind of life one may lead. Gibson says that both knowledge and acknowledgement are complementary of "one motion of the mind" (Gibson 2007, 111). What this shows is that all knowledge is expressed in how we live and act on it, and the differences in acknowledgment measure the lengths we are prepared to go in realizing the full potential of this knowledge.

3.5 The Cognitive Labor of the Work

I hope that this sketch clarifies the difference between using literary works instrumentally to give arguments or confirm hypotheses and the appreciation of the visions of our lives with concepts that we can see in Gibson's and Diamond's understanding. But what of our starting arguments? What secures the connection between the literary fictions and our world in the first place, if we do not claim that they confirm hypotheses and give arguments?

We already have enough of a story to tell about this connection. As is clear from what has been said so far, we are not generating new knowledge about the world by reading fiction because we grant that literary works are not in the business of arguing for hypotheses or giving evidence for claims. What we saw, though, is that in the process of reading fiction, we are allowing the fiction itself to work on the concepts we bring to it for the opening up of our vision of how the conceptual knowledge we already have can be hooked up to the value dimension of our embodied lives (Gibson 2007, 74). Literature does not contain or generate true propositions about the world. But this does not mean that there is no other connection to the world that literary works can have. Even if it does not refer to the facts outside of fictional worlds, story-telling is one of our ways of "giving structure to a certain conception of human experience" and it is in the business of ordering the world "by being representative of various regions of it" (Gibson 2007, 73).

But how can it be representative if we are dealing with fictions? If it does not refer, what is the connection to the world in the first place? The idea is clearly not that literature contains descriptions that refer to bits of reality. But if we accept a conception of language that is social and thoroughly conventional, we will have an opportunity to spell out the connection of language to the world prior to reference and descriptive accuracy.[14] I cannot provide a defense of this conception of language here. I can only outline what may be a way of reconciling the fictional with the actual from looking within the language.

[14]For this conception of language, see, Cavell 1969, 1979.

Very roughly, prior to the idea of assessing the representational accuracy of descriptive discourse, before, say, we can come to agreement that this thing is a table and that thing is a chair, that this remark is racist and that one is shallow, there must be public conventions and standards of representation, which we share (Gibson 2007, 63). These criteria of intelligibility have to be in place before we can get on doing anything at all in language. In Gibson's words, such criteria "provide the conditions of mutual intelligibility and so of any sort of talk at all, truth talk included" (2007, 66). According to this account of language, there are prior standards of agreement on what counts as what before we get to talking about referential accuracy and so on.

How can a culture codify and transmit standards for the mutual intelligibility of such terms as, say, 'pity' and 'devotion' or 'personhood'? It is clear that the standards for identifying when these terms apply are very complex. According to Gibson, since such standards are grounded on different visions of human life, literary tradition can be seen as one place in which such standards can be stored (2007, 77). If this is convincing, then fiction can connect to our world on the criterial level, "placing before us those narratives that hold in place and in so doing structure our understanding of large regions of cultural reality" (Gibson 2007, 74).

This is surely an incomplete presentation of a rich and exciting account of the criterial connection of literary fictions to our world. But I hope it completes a possible response to the arguments with which we started. We address the no-evidence and the no-argument arguments by showing that fiction's connection to the world is criterial. Even if it is not in the business of describing the world as the sciences do, its our-worldliness is secured on the criterial level. We read for ever-developing articulations of visions of pity, love, betrayal, personhood, and so on. Literary tradition is culture's expression of its self-understanding. Moreover, as we have seen earlier when discussing the value of detailed presentation of fictional worlds, even though literary works use knowledge we already have, the work of envisioning the concrete character of the role that this knowledge can play in living lives in the world gives an account of the cognitive value of the fictional: literature offers us

clarity of vision in regions of the lived world that other cultural resources may not cover.

To recap, what is missing in instrumentalism? The basic complaint against instrumentalism has to do with the assumption that aspects of literary fictions can be extracted from their contextual embeddedness and treated as independently available and assessable propositions. The difference, as we have seen, can be expressed by saying that the world that literary fictions present to us is a living world, not a conceptual object. The emphasis here should be placed on the work of the concretization of our conceptual knowledge. I take it that the instrumentalist extraction of bits of fictional content misrepresents the cognitive value of fictional works by presenting them as available without loss outside of the context of their concrete occurrence. This way of thinking about fiction emphasizes the work that the fictional world does with our words. Being invited to the living world of fiction is not exhausted by its aiding us in discovering what we already know and making that knowledge more precise or nuanced.

The examples of the shift in orientation to a different kind of engagement with literature given by Diamond and Gibson will serve as a guide (not an exact model, since there cannot really be a blueprint for this kind of thing) to our approach in attending to the details of the literary counterparts of fantastic occurrences.

3.6 Examples

Now that we have set the theoretical stage for thinking that fictional can be our-worldly, I want to look at some examples of fantastic occurrences in literature to reflect on the ways, in which the resources of the literary practice can be brought to bear on understanding the particular transformations that occur to the heroes of the stories we read. Engagement with the details of these examples will give us concrete instances of what these fantastic transformations can help us learn in the process of interpreting the concrete character of the fictional world.

Before we proceed, let me admit the limited scope of my claims, and the poverty of my interpretation. I cannot pretend to engage with the multiplicity of various interpre-

tations by literary critics and other scholars. Emphasizing the similarity of the chosen examples with philosophical thought experiments sidesteps their contextual situatedness in the history of literature, and the world, the authors' artistic aims and genre constraints and conventions, among other issues. My analysis also will not deal explicitly with an important question of the literary merits of these works, which may or may not have an influence on my interpretation. Claiming ignorance of this kind or purposefully ignoring these matters need not be seen simply as philosophical arrogance, I hope. Many readers of literature are in the similar situation, I would assume, without therefore thinking themselves inadequately placed to interpret what they are reading. I assume that somebody in the grip of a story may find what I say here congenial, and that is enough to get us started and hope for the acceptance of this reading sufficient for my purpose of showing the particulars of cognitive value of engaging with literary fictions along the lines of the model outlined above.

Let me also forestall the looming objection that all critical activity inevitably involves itself in some form of exactly what we have tried to avoid, namely *using* the literary merely instrumentally. Is there really such a deep difference between interpreting fiction critically while engaging with the fictional world and Carroll's interpreting it as thought experiments? I think the difference may be subtle and in some cases unclear. But to explain the difference, we have to appeal to the activity of criticism. Carroll and Kivy may wish to extract from the fictional some philosophical truth or argument, while the critic I will be playing in what follows invites others to appreciate what is in the text as meaning more than is on the paper (Gibson 2007, 142). This act of making available to thought what is on the page is the act of opening up to the reader, by another reader, what worldly significance the work may show, in the form of a complex vision of a world of the fiction (Gibson 2007, 143). It seems to me that the aims and goals of critical activity of this kind are rather different from the goals of reading literary works as thought experiments or as implying truths. While both involve articulating more (or something different) than is literally presented on the page, the goals and aims of this activity are different in each

case. Carroll's treatment of literary fictions as implicit thought experiments, or Kivy's defense of a weak version of the propositional theory of literary truth may be part of the assessment whether what is offered in the fiction is true. But the question of truth is only one of the questions that may interest us in appreciating the vision of a fictional world that is invested with worldly significance for us.[15] It may help to recall Diamond's discussion of Hume to get at the difference (Diamond 1991). To call a particular view 'shallow' is not to assess it for truth or falsity. Nor is saying that something that is 'not shallow' the same as saying that it is not false. It is to question its vision or understanding of a particular segment of human reality, to ask about the adequacy of one's stance towards that particular episode, the clarity of vision. Truth and falsity are predominantly philosophical modes of assessing discourse. The drift of Diamond's and Gibson's views was to question whether we should assume the primacy of these modes of understanding cognitive value of literary works and deny ourselves what the fictional can teach us.

3.6.1 Bodies, Social Roles, and Identification

It is predictable, when discussing fantastic transformations similar to thought-experimental scenarios in personal identity, that one start with Franz Kafka's *Metamorphosis*. The first line of the novella reads: "As Gregor Samsa awoke one morning from uneasy dreams he found himself transformed in his bed into a gigantic insect" (Kafka 1988, 67).[16] Even though Gregor is immediately aware of his new embodiment, he is neither disgusted, nor terrified, nor concerned with what happened to his body. At best, he is annoyed with the difficulty of operating this unfamiliar piece of machinery. What is he thinking about? He is worried whether he is going to make the next train, or what would happen if he had caught the train, or what the chief clerk is going to think, or whether he can pretend to be ill and keep his job. He is grumpy about other traveling sales agents who do not work as hard as he does. And so on. The first shock does not come until he hears his own voice answering his mother. Even then he is still planning to get dressed quietly and get

[15] Gibson (2007) says that before we get to truth, we must first have a vision to assess for truth (144).

[16] Kafka uses 'Ungeziepfer,' which leaves the precise anatomical nature of the beast unspecified. (Thanks to Greg Hauer for the reminder.)

going. He proceeds making reassuring speeches to his family and the chief clerk despite his appearance. This denial of what has happened takes an even more extreme form when Gregor painfully damages his jaw trying to open the door to show his readiness to carry on.

The situation is absurd: Gregor is utterly oblivious to the implications of the most incredible transformation that has happened to his body. One thing that seems clear is that taking Kafka's story to show the coherence of the psychological continuity theory of identity[17] completely misses a much richer (and philosophically relevant) vision of what is unfolding before our eyes. As the story unfolds, we learn how Gregor has incorporated the image of the family provider as his most significant self-defining feature. The strength of this identification is confirmed if we think of Gregor's reaction to the news that the family is not so bad off financially, and that in fact some of the money that Gregor had so industriously earned was stashed away without his knowledge. While we learn that Gregor's own wishes, desires, achievements and goals have been sacrificed for the well-being of the family, we also know that his self-conception over time has become interwoven so tightly with the service he provided that it obscured the most obvious of things: that he was an individual, a human being, someone to be loved and cared for. Gregor seems to be so stuck in the role he has assumed that an alternative set of reactions to his life—even when the changes are so dramatic—cannot change the logic of the story.

These details can make us painfully aware of the extraordinary degree to which one's identification with one's role shapes one's vision of and blindness to the facts. Gregor's case can be an occasion to reflect on two familiar features of our common existential predicament: the sense that we are free to open ourselves to the world and the transformative powers of life, and the equally real sense that we are conditioned not to see alternatives, that we may lose the ability to be surprised. The fantastic opening scene has its dramatic consequences because of the extraordinary juxtaposition of the most monstrous occurrence with Gregor's non-reaction. The introduction of the insect-like body

[17]This is mentioned by Baker 2000, Perry 1978, and Shoemaker 2009.

comments on Gregor's existence prior to the transformation. As we read on, one may say that in some sense little has changed with respect to Gregor's position in the family. We unquestioningly grant that Gregor turned into an insect because we are captivated by a kind of vision of the humanity/inhumanity of Gregor and his family through the lens of the comparison to an insect.[18]

Another moment that prompts reflection on identity and alienation is the episode of furniture removal from Gregor's room. His sister Grete decides that Gregor - in his new state - would like more room to enjoy himself. But Gregor's mother's remarks that this may signal to him that they do not expect him to return to his normal state, which would make him feel "forlorn." On hearing these words, Gregor realizes that "the lack of all direct human speech for the past two months together with the monotony of family life must have confused his mind, otherwise he could not account for the fact that he had quite earnestly looked forward to having his room emptied of furnishing." Gregor's new comfort would be bought, he thinks "at the price of shedding simultaneously all recollection of his human background" (Kafka 1988, 116). Gregor's fear that he is slowly fading into the insect-world is confirmed when we read Kafka's descriptions of Gregor's new habits. For example, Gregor enjoys hanging suspended from the ceiling:"it was much better than lying on the floor; one could breathe more freely; one's body swung and rocked lightly; and in the almost blissful absorption induced by this suspension it could happen to his own surprise that he let go and fell plump on the floor" (Kafka 1988, 115). This is similar to the disappearing of objects to his sight that grows weaker by day: "day by day things that were even a little way off were growing dimmer to his sight" (Kafka 1988, 112).

So, Gregor disappears into his new activities, and objects disappear for him. His world is slowly shrinking, by human standards. And so it becomes much more important for him to have some reminders, some objects for reflecting on his state, the last marks and traces of humanity, after losing which he will no longer be a person. On the one hand, then, Gregor is tempted by new possibilities for the enjoyment of his new state, of opening

[18]On the idea of metaphoric lens, see Elizabeth Camp 2009.

himself up to the new world of non-human possibilities; on the other hand, he clings to his human capacities as more defining of him. All sorts of questions arise. Is this fear of death? Of literal disappearance into the insect world? Or is this some indeterminate state of not quite knowing where the boundaries of counting as a person are? Do they lie within, or do we need the affirmation of others to count ourselves as persons?

No mention of any of the themes of alienation, loss, loneliness, and so on, is necessary in the text itself for it to be able to mean those things to its readers. What grounds our understanding that the fictional is a vision of some aspects of our own lives is the situatedness of literary practices in the the larger interpretative cultural context, which supplies the knowledge of the concepts we bring to the text. We know, for example, the feeling of disconnect from one's body, one's family, one's role in the family. But what we may not have seen so clearly before Kafka's dramatic presentation is how helpless one may be during these crises, and how deep the estrangement may go.

3.6.2 Dichotomies in Practice

Now I want to turn to an example of another prominent theme explored in literature with the aid of fantastic scenarios: the postulated distinction between mind and body, soul and flesh, and the related philosophical dualisms. Dualisms of this kind seem ready-made for the exploration of what happens at the limits of each axis: suppose I get a different body? suppose my body stays the same, but my soul/mind changes dramatically from the previous state? and so on. Let's look at Thomas Mann's retelling of the Indian legend *The Transposed Heads* as an exploration of this theme (Mann 1941).

Here is a brief summary of the fantastic occurrence. Shridaman and Nanda are best friends. Each has what the other seems to lack: Shridaman has a great mind, while Nanda has a beautiful body. Shridaman falls in love with Sita and marries her, but as time goes on Sita discovers her passion for Nanda's physical beauty. A classic love triangle gets a fantastic resolution when both Shridaman and Nanda behead themselves and Sita connects each head to a different body. Who is who now? Shridaman and Sita

are happy with his new body, while Nanda seems a bit apprehensive. A wise saint sides with Shridaman, dispensing the wisdom that "among the limbs to head the highest rank belongs" (Mann 1941, 87), meaning that Sita is married to the possessor of Shridaman's head. (It is not clear from the story whether a different answer would do just as well.) The happy couple enjoys the honeymoon. But as time goes on, Shridaman's newly acquired body and the old head adjust proportionately to the lifestyle of their owner, the result being that the original appeal of either part is neither here nor there. Sita can't stop thinking about the corresponding changes that must be occurring to Nanda. She takes off with her child to seek Nanda. Shridaman finds them and explains that the only way to put an end to the inevitable loop of their wife going back and forth is to fight a duel, in which Nanda and Shridaman remember "the double duty of giving and receiving the mortal blow," which puts an end to this cycle of dissatisfaction (Mann 1941, 113). The narrator tells us that this is the story of all marriages.

One of the caveats here is the framing device, in which the narrator inserts his running commentary. We are told that the events of married life portrayed here are "common lot" though "exaggerated and accentuated by the circumstances" (Mann 1941, 96, 99). Even without this prompting, though, we would read the legend as an exaggerated version of our own fantasies about conflict resolution in marriage by positing untenable dualisms.

When Sita puts Shridaman's head to Nanda's body and vice versa, at first this seems to be to at least her own and Shridaman's satisfacation, and indeed the two spend a wonderful honeymoon together. But as time goes on, of course, the newly acquired body adjusts to the habits and life-style of the upper-class Shridaman: the muscles lose their tone, the skin its color, the words their charm. In coordination with that, the body exerts influence on the Shridaman's head as well: he is not as sharp as he used to be, and his features are less refined. The satisfactions of artificial unification of an able body with a clever mind are short-lived. The legend shows that adding mind to a body does not function like addition in arithmetic; it also shows that being in love with a person is not reducible to being in love with a mind while also being in love with a body. Separating the

two may be useful for some purposes, but not when a couple's life together is concerned. The story shows that such dualisms are too blunt by demonstrating the false starts and hopes to solve the problem of living in marriage, with its upsets and compromises that have to be worked out by real human beings, by some miraculous cure in the shape of a new body or a new head. They are fantasies that ultimately fail to resolve the conflicts of living together.

As with the previous case, the biologically impossible fantastic details provide for the clarity of vision of how a life based on such fantasies would unfold. In the fantastic head-swap, we identify features that are structurally similar to some of our own simplifying attitudes towards living in marriage. While reading, one may be struck by the structural similarity between our own cravings to resolve the problems in our relationships by dualistic fantasizing and the legend's literal exemplification of these cravings in what each of our heroes wants. Even though the separation between mind and body tracks something deep and important, its blunt application to the case of marriage fails disastrously. Dualistic approaches to the practical problems of living do not work in this context.

Of course, no realistic situation is described here, and no predictions are being made. What we have, though, is a lens, through which we can see our own situation. And we can appreciate the significance of the mind-body interaction with the particular novel clarity that we may not have appreciated without the literary fiction.

3.6.3 Social, Ethical, and Political

Perhaps the most famous and popular cases of fantastic transformations are best known for their ability to illuminate the social and political consequences of taking one's embodiment to the limit. We will confront these themes more explicitly, even though the earlier discussions already implicitly involved the elaboration of these dimensions of our lives by means of the fictional.

Again, let's turn to Gregor Samsa's case. One of the lessons we can recover from Kafka's *Metamorphosis* is the role that the body plays in our relations with others. Even

though some philosophers assert that what we care about is the psychological continuity of others and not their embodiment, reading Kafka may bring home the complexity of the relations between embodiment and identity (or different aspects of identity) in its full concreteness.

Here we can draw on Richard Wollheim's (1984) discussion of Ovid's *Metamorphoses* to get at the problem, even though my emphasis will be different from his. In one of the stories, Io is turned into a heifer and Actaeon is turned into a deer. According to Wollheim, the story lets us appreciate what these characters either cannot do at all or cannot do in the human way. Wollheim thinks that the contrast between the inner personhood that we can imagine and the outer disguise, in which the personhood is bound to be misexpressed, is baffling. What we do not understand, Wollheim says, is how one animates the other (1984, 5–7). It is true that we can form a mental picture of a cow scratching with its hoof the letter-shaped lines in the dirt, or an image of a deer who wants to yell at the dogs chasing him, and we can superimpose on these pictures some picture of what it may be like to not be able to do things you are accustomed to. But, according to Wollheim, whole areas of our understanding are not engaged when we try to provide the rest of the picture: is the cow thinking? Does it understand language? If it does, why does it not speak to us? How does it form intentions? and so on. Wollheim claims we should treat non-human animalhood in the fantastic cases as a "disability which strikes them at their core" (1984, 7). The categories we associate with personhood—intention, action, rationality, purpose and so on—are intelligible only against the background in which they are "realized" in a particular kind of animalhood. For example, forming the intention to communicate one's name to others presupposes some repertoire of communication tools, some society in which it is accepted, and so on and so forth. Wollheim says, "if it is imaginable that there are animals, non-human animals, that are persons, then they have to be—that is, they have to be imagined to be—persons through, or in virtue of being, animals" (1984, 7). Io has to be thought of as a heifer-person, but we have no idea what

that is like. The term Wollheim reserves for these creatures are "barely persons."[19]

For Wollheim, the lesson we take from Ovid is that a person cannot simply be superadded to an animal: persons are always already human-persons, bug-persons, heifer-persons, or what have you. Wollheim's conclusion is not our concern here, but the lesson is well-taken, and it echoes what was established in the theoretical part of the chapter. If the story of Gregor is going to resonate with us, if it is truthful to how we are, the transformation cannot be absorbed without dramatic cost to one's personhood. What are these costs? Gregor's new body disrupts the kinds of interactions that maintain one's social presence for others and vice versa. Gregor's new situation impedes normal ways of expression, and similarly it compromises others' seeing their actions reciprocated in his behavior: there is no shared body of communicative signs. Even though initially there are attempts to communicate between Gregor and his sister that are expressed in her letting him *select* the food that he likes, it cannot be sustained because other ways of doing things together are thwarted.

The story shows that there could be different ways of establishing communication, but also strongly reminds us of the conditions in which expecting such re-establishing of the interaction is naive. Gregor's new embodiment, even though it does not make communication impossible, stands in the way of making such a relationship sustainable, unless both sides really try to face up to the challenge. This would involve overcoming fear, disgust, impatience, and so many other things, and it may be impossible despite efforts to overcome these deeply ingrained reactions. The details of the story reveal, then that the sustainability of some of our ways of being with others depends on the willingness to communicate, the familiarity of our bodily form, and the availability of some ways of communication, and also that there are limits to our powers to influence these constraints. This has repercussions for one's treatment at the hands of others and may even result in denials of one's history.[20]

[19] See David Cockburn, *Human Beings*, 1991, pursuing the same theme.

[20] In a dramatic change of attitude towards her brother, Grete announces that to solve 'the problem' of Gregor, they must get rid of the idea that this creature is Gregor because Gregor would have spared

Perhaps the example that explores even further the issue of the sustainability of our practices towards others, is Karel Capek's *The War with the Newts*. Capek's novel—even though it is most immediately a poignant satire of Capek's contemporaries (the satire speaks to our times as well[21])—is nevertheless very fruitful in exploring which practices and attitudes are sustainable given the facts. Starting from a small colony on a remote island, newts come to dominate the planet and eventually compete with the humans for space, eventually threatening their existence. Between these two points, both the newts and the humans undergo a very serious transformation. Capek's narrative brings out various interrelated aspects of both integrating the newts into human society and of organizing their own.

Here are some examples of various conceptual entanglements that can come out in our engaging with the details of the story. For example, if newts can speak the language and reason, and do so in some cases better than humans, their treatment as mere beasts (including performing horrible experiments on them) is clearly immoral. If they submit articles for scientific publication, denying their intelligence is hypocritical. If you sell them weapons, you should not be surprised that they will use them to defend themselves. If humans employ them and they constitute the largest work force, they need universal education, will demand improved working conditions, and will be represented in court by (human) lawyers. And so on. These points are meant to demonstrate the "hardness of the social 'must,'" if you will. What emerges is that the fantastic allows us to explore the real patterns of dependence between various aspects of social existence, like the relation between labor and rights, intelligence and respect, the shared existence with another species and the change in attitudes we can observe as time goes on, and so on.

Even though none of the occurrences here are actual and no predictions are being made

them all this suffering (Kafka 1988, 134). So, in terms of practice, one of the "solutions" may just be to simply decide that the giant insect in the room is not Gregor. This could not have been possible—or at least it would have been very unlikely—if Gregor still maintained some human features. But an insect's body is so foreign to us that we can in fact decide that it is impossible for a family member to continue as one.

[21]Marya Schechtman reports seeing a Next Theater, Chicago production of Capek that incorporated the Mexican Gulf oil spill of 2010. The novel itself dates to 1936, which, by all accounts, was a fertile year for satire.

about how such events could possibly unfold, the connections between the considerations of language acquisition, intelligence, employment, rights, and justice are actual at the criterial level. Exploring the generation, molding, and understanding of the newt society—in evolutionary, ethical, pragmatic, economic, or whatever, senses—tells us very much about ourselves and allows us to occupy a standpoint from which we can criticize our own ways. Once we have noticed these similarities, we may be in a better position to explore and compare the different options that simultaneous attention to both the fantastic and the actual affords. I.e., besides being engaged in a descriptive project of imaginative construction, we are applying various evaluative and normative criteria to the fantastic and the actual social worlds, sharpening our understanding of these criteria.

3.6.4 Making It More Abstract

At this point, I want to turn to an example that presents the fantastic as an opportunity to pose questions of a more abstract level: how should we think of the encounter with some radically new possibilities? On what kind of resources for understanding others and ourselves can we rely? This particular example may not quite fit the line of thinking I have been developing because one may object that it is a kind of hybrid between literary fiction and philosophical concerns. Still, I include it here because I think it can be interpreted alongside the examples I already presented.

The difficulty of making evaluative judgments about the relation between physical appearance, biological constitution, and person-related practices is explored in Stanislaw Lem's *Solaris*—familiar to many by Tarkowski's and Soderberg's film adaptations. Here is a very brief recap.

Kris Kelvin is sent to a remote space station near the planet Solaris to investigate strange signals from it. When he arrives at the station, he discovers that the remaining crew of the station all have guests that are human in appearance but completely foreign in material constitution. (For example, they have the ability to regenerate after damage that would be fatal to humans.) The visitors seem to be dependent, in some way, on the

visitee's memories. Kelvin's visitor takes the form of his dead wife. One of the theories about visitors' origins has to do with the activity of the planet Solaris, according to which the planet exhibits a form of consciousness that is foreign to us.

Among many other issues of philosophical significance, it explores the question of which kind of stance should be taken towards the "visitors" that show up at a remote space station near the planet Solaris. The visitors look like humans, have memories, and so on, but are not biologically human. For one thing, their molecular constitution is foreign to ours, which makes their tissue regenerate (if such terms are appropriate here). The remaining crew of the station develops various ways of interacting with the visitors and understanding the phenomenon, and thereby may be taken to exhibit various criteria each of them takes to be defining of the kind of attitude a person should take to his visitor. One of the crew members, Sartorius, in the name of science, fearlessly experiments on them despite their pain-behavior; to him, they cannot be persons or have human moral status because they are not biologically human. Another one, Snaut contemplates the more abstract question of what this could mean for humans' self-understanding, but is not averse to cruel measures. Kelvin, by living together with his visitor, falls in love with her, even though at the beginning he is rather cruel to his visitor. Importantly, this change of orientation is not based on his having access to more factual knowledge than is available to either one of his crew mates. Rather, it is based on a certain living orientation to them, or certain emphases given to some features of reality rather than others. For our purposes, the novel shows that even when all the facts are in, as it were, some decisions have to be made about which criterion of identity, or which features of reality, to accept as definitional of persons or meriting person-like responses: biology? Our own reactions? Their psychology? By engaging with the fictional world, in which these different approaches are elaborated, and being exposed to their results—be it madness or becoming more and more insensitive to others, or whatever—we can appreciate different visions of humanity elaborated with us.

In *Solaris*, perhaps because of the more abstract philosophical concerns that Lem was

exploring, the question of sustainable social practices is not explored further than the positing of the problem I have sketched. But we can imagine the story continuing [22] in different ways, and we can also imagine that some of these ways will strike us, as some already do,—by whatever criteria we deem appropriate—as better and worse, plausible or implausible, ways to go. What is important to notice at this point in the discussion is the connection between the decisions concerning the appropriate treatment of visitors and the considerations of personal identity—biological, psychological, etc.—that we will give as reasons to justify or call into question a certain attitude towards ourselves and others.

This concludes my brief sampler of examples—dramatic bodily transformations, head-swaps, interactions with non-human persons, and so on—of what we called the literary counterparts of thought experiments in personal identity because they in various ways manipulate aspects of our existence that normally go together. I hope that my brief and amateurish excursions into the details of these fantastic stories can illuminate the powerful tool of imagination at our disposal. The fictional is our-worldly in this sense: it can articulate multiple interrelated aspects of our culture's self-understanding by inviting us to engage with fictional worlds.

3.7 Conclusion

According to Wilkes's objection discussed in the previous chapter, philosophical thought experiments are inconclusive because without the full relevant background they cannot begin to address metaphysical questions of identity. Literary examples of fantastic transformations may at first sight seem to take care of this worry—their elaborate background can serve to make them into better conceptual thought experiments than the ones used by philosophers. We saw, however, that thinking of literary works as elaborate arguments or implying certain hypotheses instrumentalizes literary fictions, and that there are significant difficulties with and worries about this project. I have appealed to the work of

[22] I would like to credit Marya Schechtman for this suggestion.

Martha Nussbaum, Cora Diamond, and, most explicitly, John Gibson, to make the case for understanding the value of engaging with literary fiction differently. This was not to deny the value of traditional philosophical tools like argumentation and reasoning, of course, but to suggest looking for resources that are internal to engagements with fictional literature as such (*not* as containing implied arguments or hypotheses). As I said in the introduction, the point of looking at literary fictions was not to read off them the answers to the traditional metaphysical questions of identity, but to explore the elements that make fictions cognitively significant without making them into arguments.[23] The full effect of looking at literary examples for philosophy is not going to come to the surface until Chapter Five where I bring the lessons from Chapters Three and Four to bear on thought experiments in personal identity.

Even though each fantastic fiction we looked at is different, we have tried to identify some of the unifying themes that are explored in many of them. Because of the surface similarity of the transformations invoked in such fictions with those in philosophical thought experiments, there is a thematic connection between them. The themes explored in literary fictions reflect identity-related practical questions in the contexts of lives that either undergo an extraordinary transformation or witness such a transformation in others. These themes, in one way or another, explore the role of body and oneness in self-understanding or understanding our social arrangements, in both first- and third-personal terms, or they explore questions of the limits of recognition of someone as deserving this or that treatment, based on the change in her body or mind, along with exploring the sustainability of social arrangements given various possible transgressions that violate normal expectations of what may or may not happen to one's body or mind. These concerns are very natural to creatures who are leading lives that involve various kinds of changes, one of the most significant of which is the process of growing up, getting older, and dying. The terms of our self-understanding presuppose this kind of paradigmatic trajectory that

[23]The rich resources of our literary practices involve different tools that can aid us in engagement with the fictional worlds, such as metaphoric comparison, dramatic juxtaposition, cogitation on a theme, pretense and so on and so forth. A more detailed examination of these tools is beyond the scope of this project, however.

a life takes, along with paradigms of acceptable changes within a normal life trajectory. Literary explorations of various deviations from this expected trajectory—what happens at the margins of the expected—give us visions of possible life configurations and, by doing that, bring into sharper focus and develop the understanding of the normal trajectories as well.[24]

Now, the answers we get from such explorations cannot be seen as accurate descriptions of how this or that scenario will in fact turn out. In each case we discussed, we can imagine possibilities for continuing the story differently than did the the story's author, and this is part of our interest in reading fiction. In addition, there is no reason to think that, presented with such cases, we cannot have conflicting visions of what one could or should do, or even what it is that we are reading about and its significance for us.[25] What is crucial for my methodological goals is the way in which literary fictions exhibit complicated interactions between our practices (including, of course, our evaluative stances and ethical concerns) and the material aspects of our being—embodiment, uniqueness, fragility, and so on. Engaging with literary fictions shows that these intersections turn out to be immensely complicated, and we quickly realize that our views on the persistence conditions of objects like ourselves are not quite independent of the social and ethical dimensions of our lives. That is, one of the significant lessons of thinking about literary fictions is the idea that material aspects of our being, including our persistence conditions, may be dependent on the aspects of our lives that are traditionally thought to be secondary to the discussion of metaphysical facts. So, considerations of practice and value may be directly relevant to considerations of facts of the matter when it comes to discussions of persons. Thus, the world-making of literary fiction can expose the limitations of the

[24]One might think that I am suggesting that, because of the thematic affinity between the literary counterparts of the philosophical thought experiments and the thought experiments themselves, the suggestion is dangerously close to the one I have been denying: that literary fictions are better thought experiments because they are more detailed. I hope, though, that I have been explicit enough to distance myself from the view that suggests that we should read off the literary fictions the answers we seek to the philosophical questions of personal identity. In part, the selection of the examples was motivated by my earlier instrumentalist approach at the early dissertation stages. I am now a non-instrumentalist, but the examples still work.

[25]I am grateful to Matthew Pianalto for talking to me about this issue.

current method of thought experiments as inadequately abstracted from the phenomena of actual life which they are presumed to illuminate. It also shows that rather than directly addressing questions of survival, literary fictions try to present a coherent world in which particular fantastic circumstance can hold. This shift of orientation will be most significant when we return to philosophical thought experiments in personal identity in Chapter Five.

This contextual dependence implies an open-endedness in our thinking about such cases, and we should admit that this may be uncomfortable if we assume that our goal in thought experimentation is to untangle conceptual puzzles and to directly address the questions of identity. But now that we have shifted to viewing thought experiments as being exploratory of the complicated material and conceptual background assumed in asking the questions of personal identity, their usefulness does not have to be defined by its ability to answer questions posed at the outset. Instead, its more important achievement is in showing us that we can refuse, modify, or understand differently such questions and answers to them.

As we saw from the discussion of John Gibson's work, we have to admit that literary fictions must presuppose some kind of background knowledge, and in this sense it may seem that they are somewhat secondary because of their reliance on a more general cultural understanding. The worry above suggests that the knowledge we come to recognize as we read is already present there *prior* to the engagement with the literary fiction, and so what is shown in our judging that some works "ring true" while others "merely make things up" must be independent of the engagement with the fiction itself. But is this the only option for thinking of the relation between literary fictions and knowledge, namely that antecedently present knowledge is revealed by literary excavations?[26] I think that the relation between knowledge and fiction can be thought of as being mutually supportive: what we recognize as "ringing true" and being "appropriate" in a life of a given character is not independent of the relation between the reader and the fiction itself; i.e., the work

[26]I think this view may bring us back to an instrumentalist understanding of the usefulness of literary fiction for philosophy.

done by the engagement with a given fiction is necessary for us to understand what it is that we actually know. That is, when we tell good fiction from bad fiction, we do not have the ranking criteria ahead of the process of engagement; rather, the criteria emerge in the act of reading the work of fiction. So, in some sense, knowing how to go on with our concepts may not be available without engaging with fiction.[27]

[27]Of course, this does not mean that such criteria are beyond critical examination.

Chapter 4

Thought Experiments in Bioethics

4.1 Introduction

In the previous chapter, we established that literary fictions are cognitively significant because they are representative of the complex understanding of our cultural practices: they reveal the axiological dimension of our life with concepts. While literature must depend on what we already know, this does not imply that the knowledge we gain is trivial. Fiction's power lies in its ability to work on the concepts and knowledge we already possess. According to this view, fiction's role lies in showing the connection of the conceptual knowledge to the axiological dimension of our lives. Additionally, the full understanding of knowledge and rationality requires thinking through the entanglements that particular concepts have with our world as an object of our concern, and how such understanding reveals possible interactions between different dimensions of our lives, especially the connection between the material aspects of our being and the concepts and values we have.

Most of my commentary focused on ethical and socio-political dimensions of our living together exposed by fiction. Thus, we can see our own alienation from what our work and family through the lens of Kafka's novella; we can question the simplistic separation of the mind and body by reading Mann's novella. However, one might wonder what such discussions have to do with metaphysics of identity. It may seem that what I said about

the cognitive value of fiction is far too general and requires additional resources to narrow it down. Additionally, I may already be guilty of confusing the power of literary fictions to illuminate questions of value with something that is less clear, namely its ability to address the metaphysical questions. While I will not explicitly deal with the complicated question of the interaction between metaphysics of identity and our practical concerns until later in the thesis, let me provide the initial justification for turning our attention to bioethics to address the first part of the complaint.

Most generally, looking at the role of fantastic scenarios in bioethics in this chapter presents a narrowing of the general approach described in Chapter Three to more particular questions about the interaction between our practical concerns and the tangible embodied entity of our personhood. Of course, I cannot deny that literary fictions are capable of bringing up these more narrow questions, and ultimately I think the boundaries between the two genres in some cases may be blurry. Nevertheless, I think it is helpful at this stage to focus directly on the power of imagination to engage such more narrowly defined questions in a more constrained setting.

Second, looking at bioethics is justified because it presents a middle-ground between Wilkes's approach to scientific thought experiments and the full imaginative engagement of fiction. This feature is significant for our purposes because it introduces the idea of guided and constrained imaginative exploration. On the one hand, bioethical reflection has to be explicitly set in the context of the current and projected scientific knowledge. On the other, it cannot ignore the potential of technology (including far-fetched options) to question and overthrow some aspects of our self-understanding. Thus, from the methodological standpoint, looking at bioethics may contain insights for constrained uses of imagination in personal identity.

The third reason is that there is a significant difficulty in navigating between the cognitive import of the stories for the purposes of providing a model for philosophical thought experiments and the aesthetic aims of the author, along with the contextual situatedness of a given literary fiction in our cultural history. Thus, full understanding of

the artistic purposes may come in conflict with my readings of the works I have chosen, and there may be serious problems with justifying the abstraction of those readings from the context of the production of the works. Turning our attention to bioethics provides us with a context, in which this particular problem is not as severe. I presume that the authors of the fantastic thought experiments in bioethics do not explicitly pursue aesthetic goals but are driven by the pragmatic goals of advancing a particular argument in defense of some theoretical point, or an actual policy-making decision, or a decision about a particular patient's case, or with the pedagogical goal in mind.

In addition, the remarks in the previous chapter may seem a bit too abstract to deal with some questions,about the influence that the visions of far-fetched fantastic possibilities may have on our self-understanding. Thus, Kafka's novella is more likely to be seen as a metaphoric lens for understanding the issues of alienation or oppression rather than as a vehicle for understanding the issues connected to embodiment. Bioethical discussions, however, are more intimately connected to the technological possibilities of changing the very basic components of our embodiment and thus our ability to control certain aspects of our lives in ways we could not envision before.[1] New ways to interfere with and direct the course of a given life open negotiation-space about the questions of what exactly constitutes a given life or what it is to be a person. The technological advances in our ability to manipulate life should prompt the new sense of urgency about the ethical norms governing our conduct. Such close interaction between practical concerns (including ethics) and the question of the entity (metaphysics) is currently at the center of some methodological discussions in the philosophy of personal identity. Thus, it seems natural to think that the discussion of bioethics may contain valuable methodological insights for the metaphysics of personal identity which in the mainstream shares the assumption of the tight connection between personal identity and practical concerns. (I directly confront these issues in Chapters Six and Seven.)

[1]Do we sustain the brain activity with an artificial heart-lung machine? Do we order dialysis to keep the person alive? Do we abort the fetus if it carries a certain gene? These questions are prompted by our ability to know and manage some parts of our embodiment that were not negotiable before.

Lastly, fantastic possibilities discussed in bioethics urge us to think that imagined scenarios, including literary fictions, are "closer to home" than one might think. The context of bioethics makes more perspicuous the idea of dynamic boundaries of concept application, which are neither arbitrary nor fixed in stone as technology marches forward.

I proceed as follows. I first give a sample of different cases from bioethics in order to give a flavor of the fantastic element present in them. One might think that Wilkes's model is applicable to such cases. Section 3 is a reminder of why Wilkes thinks that the actual cases are better than the fantastic ones. In section 4, I discuss some cases that follow Wilkes's model. In Section 5, I discuss some proposals to understand the import of the fantastic elements as being similar in spirit to the scientific model. As we will see, does not capture all aspects of our appeal to imagination in these contexts. In Section 6, I discuss the proposal to diversify the functions that thought experiments may perform. I think the move is still insufficient to account for all elements at work in imagination (Section 7). In some cases, our imaginative exercises guide our vision and assessment of different future possibilities: they show that some of the possible futures are more desirable than others, or more stable, or whatever. In the process of imaginative engagement, we clarify to ourselves the dimensions of our lives that may be easier to change than others or show us more desirable directions of such changes. I do not suggest that what we can imagine determines or predicts the future. But what is revealed by imagination in the form of our reactions of endorsement or resistance to this or that possibility is not idle either. Since practice is inherently dynamic, the future possibilities are continuous with the present, and imagination may guide future extensions of our practices. I substantiate these abstract remarks with an example from Hilde Lindemann Nelson (Section 8). Section 9 revisits some of the earlier examples to show the idea of constrained imagination at work.

4.2 A Sample of Cases

Bioethicists operate at the intersection of multiple practical issues: policy and individual decision-making, family rights and obligations, medical practice, financial institutions,

technological realities and future possibilities. It is not surprising that an applied field like bioethics revolves around cases. Some of these cases are actual court cases or public statements by individuals of their stories.[2] Others are "doctored" to various degrees to preserve anonymity.[3] Yet others are made up to provide medical professionals with "reflection scenarios." In this chapter we will be concerned with the more fantastic variety of the latter sort, ranging from the ones that may soon become reality to those the possibility of which is not easy to determine.[4]

Consider the following cases:

> You wake up in the morning and find yourself in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you , and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital tells you, "Look, we're sorry the Society of Music Lovers did this to you - we would never have permitted it if we had known. But still, they did it and the violinist is now plugged into you. To unplug you would be to kill him. But never mind, it's only for nine months. By then we will have recovered from his ailment, and can safely be unplugged from you." (Thomson 1999, 88)

[2]Some famous cases are the Dax case, and the Quinlan case, for example.

[3]There are interesting discussions of fictionality of such cases. This may bring them closer to the literary explorations of the earlier chapter. Even though there are disputes on what the fictionalizing does or whether it is possible to treat these as actual when so many of the parameters from the actual cases have changed—gender, age, kind of disease, family situation, they do not yet involve impossibilities. See Arras 1991, 1994, 1997.

[4]The situation is more complicated, of course. First and foremost, the presence of the variety of the cases I mentioned, demonstrates the aspirations of bioethics as an applied field by connecting actual problems that arise in daily medical and institutional practice with the theoretical achievements in general and applied ethics. Second, many of the cases—like Quinlan case, or the Tuskegee syphilis study, etc.—form the historical background of various discussions in bioethics, like informed consent, patient's autonomy, etc. (This connection is particularly strong for the devotees of the casuistry school of bioethics, who argue that reasoning by analogy from the lineage of historically well-established cases should be preferred as the methodological approach to the more theoretically driven approaches like principlism or the top-down approach. (See Jonsen and Toulmin 1990, Arras 1991, 1994, 1997; Beauchamp and Childress 2008.) Since bioethical reflection is located on the intersection of practical problem-solving and policy-making, these cases are not just mere illustrations, but are constitutive of the past thinking on the subject. Third, cases give an opportunity to properly address the theoretical issues of contextualization that is often absent from the more abstract discussions of theoretical ethics. For example, in making decisions about the proper exercise of paternalism in a given case, it may be useful to consider not simply the legal age of the person, over whom the authority may be exercised, but her level of education, income, psychological profile, and so on. This range is by no means exhaustive but it should capture the spirit of the enterprise.

> A disaffected member of what the media refer to as a religious cult announces that the group is attempting to implement its vision of the good society by "mass producing" human embryos cloned from the group's leaders. He claims that the group has its own genetics lab and hopes to adapt for use on humans techniques for cloning embryos commonly employed in the commercial production of animals. Several members of Congress express outrage and urge that the government take action against the religious group. A spokesperson for the American Civil Liberties Union says that if we value reproductive freedom and freedom of religion, we must respect the right of religious communities to attempt to transmit their beliefs and way of life to future generations, whether by the traditional methods of teaching and indoctrination or by the application of genetic technology. (Buchanan et al. 2001, 2)

> Imagine a superdrug is developed for altering mood, which has no side-effects involving any risk or impairment to health, and which is in no way addictive. It does not lead to any blurring of awareness or to lethargy. Its effect, on the contrary, is to make people more alert and invigorated. Although it induces a good mood and a sense of well-being under normal conditions, its effect is not so overriding as to rule out different emotional responses where they are appropriate to the person's beliefs about what happens. It induces no tendency to social quietism. ... [it] is never used to influence others' behavior, but is only used by people to affect their own mood. . Everyone in the community is fully informed about its effects, and it it s only taken knowingly and voluntarily. (Glover 1984, 76)

> Imagine a world, produced by utopian adjusters, which too, account of the desirability of autonomy and variety. In one place is Sisyphus, contentedly, and of his own choice, rolling stones. Nearby is someone else, hopping about in complicated patterns. There are thousands of such people. They are spending their lives building towers out of marzipan, writing articles about meaning, knitting huge maps of the moon. They are all autonomous, contented and different. (Glover 1984, 163)

Each of these scenarios is used for different purposes. Judith Jarvis Thomson's violinist case is given in the context of the discussion of permissibility of abortion. Allen Buchanan, Dan Brock, Norman Daniels and Daniel Wickler (2001) are reflecting on the foundations of the values that a just society with radical powers of genetic interference should have. Jonathan Glover (1984), among other issues, explores the depth of our objections to mood enhancement drugs. The authors of these examples may also differ in how they justify the use of the fantastic scenarios.

The fantastic cases above are thought to be continuous with the actual cases, such as mood-enhancing drugs, cloning, and organ transplantation, both in their content and in how they are used as pedagogical tools.[5] What matters in these cases is the new power of technology to control and direct certain aspects of our being. Even though we do not know whether and how certain technological breakthroughs will affect our social arrangements, we can speculate about such possible developments.

4.3 A Reminder about Wilkes

In the second chapter, we discussed Wilkes's dilemma. If we hold the relevant background of our world fixed, the scenarios described in thought experiments are impossible; if we stipulate a different background, then we do not know what to say about these situations (Wilkes 1988, 46). Concluding that thought experiments are not fruitful, Wilkes thinks that actual puzzling cases are difficult enough and provide plenty of opportunities for us to sharpen our understanding. How do these actual cases sharpen our views?

Consider the famous case of Laura Beauchamp, who displayed different, independent, and at times conflicting personalities (Wilkes 1988, 109–128).[6] Hard cases like MPD put pressure on our concepts. Wilkes says that in MPD "all the facts are in, or can in principle be collected (albeit they need careful handling, description, and interpretation)" (1988, 128). Based on such facts, Wilkes says, we can ask what people (including the medical community, first and foremost) say. According to Wilkes, actual cases are better than the fantastic ones because we have a clear conception of the norms that warrant and guide application of our concepts like 'person,' and we can determine the source of conflicting intuitions. On the one hand, in such cases we observe different disconnected strands of psychological continuity: there are separate disconnected unities we witness, we want to say. On the other hand, there is normative pressure to treat the patient as a single

[5]Of course, there may be intractable incoherences in thinking that genetic clones can threaten individuality, and that making each individual fully individually autonomous and happy may be as impossible as cerebrum transplantation (Sober 2001).

[6]I will use her terminology even though it is outdated. Currently 'Dissociative Identity Disorder' (DID) is used to describe such cases.

individual despite reasons for plurality. Once the conflict is identified, Wilkes's proposal is to give up on one of the background "conditions of personhood" that generate it: what we see is that demands of strong unity of consciousness as a precondition of personhood cannot really be sustained in anomalous cases. Thus, the positive result of reflection on actual cases is that they should prompt us to reflect systematically on various criteria and situations, in which we use words like 'person' or 'personhood' (Wilkes 1988, 158).

There is a lot to say about Wilkes's argument, but only some details are important for our purposes. At this point, note that Wilkes's argument depends on "careful handling, description, and interpretation" of the facts of the actual hard cases. Wilkes would like to maintain that the matters of interpretation and description of the hard cases themselves do not have to appeal to imaginative projection, fanciful hypothesizing, and so on. However, as I argue in Sections 6, 7, and 8, the relation between imagination and "facts of the matter" is not as straightforward as Wilkes may think.

Before we get to it, however, let me devote some time to some thought experiments in bioethics context that actually seem to fit Wilkes's model because in them the postulation of the fantastic variable is not destructively relevant to the thought experiment. Judith Jarvis Thomson's violinist case (discussed in the next section) is one such thought experiment that may be interpreted along the lines outlined by Wilkes's in the second chapter.

4.4 Thomson's Violinist Case

Recall Thomson's violinist case from our sample. Thomson (1999) provides the case in the context of showing that even granting the status of a person to the fetus is not sufficient to establish that aborting an unwanted fetus is not permissible. The case is artificially constructed to exhibit a relation between two persons that is analogous to pregnancy.[7] In this case, what serves as proxy for the scientific theories are our commonsense convictions that are presumably widely shared. According to Thomson, most of us would agree what

[7]Nobody else can filter the blood for the connected person, premature unplugging will kill that person, you are connected for nine months, you did not consent to be "hooked up", and so on.

to say in the violinist case, and the argument goes as follows. The standard framework of the debate—the relation between the prospective mother and the fetus, the sentiments we have about the beginnings of life, religious views we may hold, views on person / potential person / something-with-future-like-ours, and so on—are explicitly ignored. Assuming that fetuses are persons, you have to apply the rules to them that apply to all persons. So, either you have to say that you cannot unplug from the violinist, or you have to say that fetuses are not persons, or you have to say that fetuses are special persons. If you say that fetuses are special persons, then it is not their status as persons that plays the crucial role in our arguments against abortion. If you say that you cannot unplug from the violinist, you go against a very strong and wide-spread intuition that it is fine to unplug in this case.

Presumably, what is doing the work in our analogical reasoning here is the presence of the shared background of what we find intuitively permissible. Here, the figure of the violinist is used to focus our attention just on the relation holding between two persons, abstracting from other details. Without the assumed framework supporting our widely shared judgment of permissibility of unplugging in this case, the case would be a mere curiosity. This is not to say, of course, that all parties share this framework or that the analogy is good. But there is a structural fit of the violinist case and Wilkes's description of the desiderata of successful thought experiment. That is, whatever the questions one may have about the analogy itself, they are, presumably, not about the technological possibility of hooking people up to each other.[8]

4.5 Glover's "Controls"

Can other fantastic cases be understood along the lines we just used to approach the violinist case? Jonathan Glover's (1984) remarks on the methodology of thought experiments in bioethics seem to be fitting Wilkes's model.

Glover (1984) explicitly discusses the methodology of thought experiments—the method

[8] As we go on, we may in fact question the idea that the technological possibility is not relevant here, but I think that what I say here is not controversial.

he heavily utilizes. Some of these are rather fantastic, if you recall his examples from the beginning of this chapter. These cases are not even the most fantastic of the ones he uses. Others include the possibility of transparent minds being read by others, direct brain manipulation for control of behavior, varieties of experience machines fashioned after Nozick's famous example (1977), and so on. Glover confesses that many have objected to the use of his thought experiments on the grounds that they are simply too far removed from the world of our experience, too unrealistic, to teach us anything about our world (1984, 17). Here is why he thinks they are useful despite being fantastic.

First, according to him, "[t]he complexity of practical detail, so essential to a decision in a particular context, has a softening and blurring effect when we are trying to think of what our priorities are in general" (Glover 1984, 17) Second, in 'moderate' cases, we are aware of being on a slippery slope of a spectrum with some terrible end, but "it lurks unconfronted as the thought of some nameless horror," until we confront it explicitly (Glover 1984, 17). And third, in discussions of moderate cases, "we are hardly ever able to choose between kinds of life in general... We are limited to piecemeal decisions" (Glover 1984, 17).

Looking back at the superdrug case, let's work out what the actual proposal is, given Glover's own remarks. The first two points—sharpening our focus on the cases and explicitly questioning our deep assumptions—may go hand in hand and seem to work according to Wilkes's model. The superdrug case is the end of a series of successive modifications of the previous versions of the drug. For example, to avoid the objection that such global soma-like treatment can be used by governments to preclude social unrest, it is stipulated that the drug is not lethargy-inducing.[9] To confront the objection that "happy drugs" threaten autonomy and can cause coercion, it is stipulated that the users possess all relevant information and never coerce others. Suppose we take care of all these objections. According to Glover, some people still find the drug objectionable on the grounds of its "unnaturalness" (1984, 76).[10]

[9]Glover mentions Huxley's *Brave New World*, and I will return to this connection.

[10]The same kind of step-by-step sharpening of the case occurs in the "human zoo" from our sample.

According to Glover, by eliminating the blurring details of the objections to the earlier cases, and focusing on the objection from "unnaturalness", we can press for a clarification of this worry. One of Glover's more general conclusions is that there may be nothing *justifying* this objection except for the prejudice in favor of the entrenched conventional understanding. So, it is rational to conclude that careful progress in making people happier by chemical means is permissible unless we have an explanation of the sharp line between such means and the presumably traditional and unobjectionable practices of mind-alteration, such as prayer, reading poetry, doing math, schooling children, watching a funny movie, and so on (1984). (This reasoning can be applied to the genetic manipulation for the purpose of conditioning an individual to fit society.) As a payoff, without the thought experiment the charge of "unnaturalness" raised against superdrugw or other direct manipulations of our bodies—often going by name of "playing god" or some such idea (e.g., Sandel 2004)—would not have been brought into sharp relief and shown to be problematic.[11]

Notice that Glover's procedure seems to be analogous both to Wilkes's and Thomson's approaches. We manipulate the impact of the drug on both individual and social levels to check whether we would find the drug objectionable, while assuming that our values stay the same. Since the possible worlds we are discussing can be assumed to be close enough to the actual world, our intuitive responses can be assumed to be reliable. Glover's claim is that "bare and abstract" descriptions are a way to control for biases that would be generated by the more realistic accounts (1984, 131). The language of "controlling for biases" is intended to make the analogy with science explicit. By fixing different features of the background for our purposes, we are able to check the changes in our reactions and discover our more fundamental values and concerns that are otherwise buried under the distracting details of the actual world. Glover also provides a psychological explanation of

It describes an alternative world, to which the objections on the grounds of loss of autonomy and variety supposedly do not apply, and so those who think there is something objectionable in the picture as presented have to reexamine the source of their dissatisfaction by imagining this radical alternative to our world.

[11] DeGrazia 2005 is another source for a more contemporary discussion of these issues.

the power of abstract scenarios. According to him, they often threaten our value systems and thus bring to the surface the values that have been deeply buried underneath those that are more explicitly endorsed.[12]

Glover's position is attractive: we take a phenomenon, dissect it, take some parts out, input it into our thinking, and consider the resulting judgments.[13] However, there are problems in assuming that we can "control" for our reactions in the way Glover suggests. This has to do with holistic nature of our values. Let me elaborate.

I think Glover's comparison of bioethical cases with those in sciences is overstated. Only by a very rough approximation what is going on here is like controlled experimentation in science. It is not controversial, of course, to discuss the values of autonomy, deep satisfaction of living with others, authenticity, and so on, separately from other values like those of justice and not harming others. But thinking that our reactions to alternative pictures of the worlds in which these features have been "adjusted for" one-by-one are also neatly adjusted to calibrate for the world with fewer and fewer objectionable features, presupposes that our intuitive responses can be clearly correlated with discrete features of those worlds. I think, however, that such (arithmetical) adding of values and/or of reactions to their presence or absence are a simplification of a much messier picture. The clusters of our attitudes and values form dynamic wholes which are subject to all sorts of local modifications and conditions that it would be hard to know where to start separating them. At the same time, methodologically, the task of controlling for each of the possible influences on our reporting our reactions to this or that thought experiment is very difficult. The judgments based on thought experiments are too difficult to assimilate to the model of science, because they are subject to so many ceteris paribus clauses that their systematization would be counterproductive.[14]

[12]For example, we often confront the issues having to do with fairly abstract values of justice, fairness, impartiality. On the other hand, the values of self-expression, of contact with others and of variety of experiences we have with different people, the value of experiencing growth and maturity of our responses to the world, and the value of deep satisfaction of accessing the learning experience itself do not usually come to the surface in our daily lives but are held implicitly. Glover thinks such values are harder to discern and articulate (Glover 1984, 132).

[13]On the surface, this looks like the standard method of variation familiar from J.S. Mill.

[14]Wilkes's work on misunderstandings of what folk psychological generalizations can and cannot do is

Applying this to Glover's discussion of the "unnaturalness" objection, then, we should question whether the sources of our resistance to the superdrug can be so easily unmasked. Instead, what seems right is that our objections are only as deep as our practice goes and what is called "unnatural" may be a function of various risks and benefits (and their dynamic interactions in different cases) associated with not fitting the norm. In order to have the contrast between naturalness and its absence we have to presuppose some stable habits of living, judgment, dispositions, and so on. But in the course of gradual thought-experimental adjustment for this or that, these features of the background are subject to step-by-step adjustment. But then the idea of a "control", similar to the one operative in science, is misplaced because the variables we are fixing—social risks, coercion, no side-effects, etc.—are intertwined.

I admit that Glover is right that often unjustified prejudices are hidden behind the talk of unnaturalness. I am prepared to accept the idea that we can get a rough correlation between our reactions and the gradually adjusted social effects of the superdrug. So, suppose we accept that we can get a rough idea of "control" for some coarser grain of focus. Two things need to be said, however. Glover's methodology is general and should apply to his other examples, like mind control, experience machines, human zoos, and so on. In some of these cases, however, our intuitions cannot be as reliable as they may be in the superdrug case because we simply do not know what the background of Glover's perfected worlds—by means of stipulating away of the features—will be. Thus, Wilkes's objection applies to many of Glover's more far-fetched thought experiments. More importantly, since Glover is interested in contemplating our technology-driven practices and their future, we have to picture the role of imagination differently from simply giving us cases that test our existing ethical theories (as they do in Thomson) because the interaction between the fantastic technology and our norms is often more complicated than it is presented in the picture we are discussing.

I earlier said that the superdrug case is not quite like Thomson's case in which the

instructive here (1981, 1986, 1991).

use of the fantastic elements is not destructively relevant to the conclusions drawn from it. This brings us to Glover's third reason for using thought experiments, namely that they allow us to contemplate different "kinds of life" in general. The superdrug case, for example, in addition to clarifying our current value system, prepares us for the alternative possibilities of viewing the difference (or the absence thereof) between "direct" and "indirect" manipulation of the brain to create certain moods. This interpretation of thought experiments was not playing a prominent role in our discussion up to now. However, radical abstraction from the messy details can reconfigure our orientation towards deep and important distinctions, giving us resources to distance ourselves even from the currently assumed framework of values. For example, what difference does it really make if I learn mnemonic techniques or if I take the drug to improve my memory? Why is one less "natural" than another? We cannot simply rule out alternative futures, in which direct chemical boosts to the brain are an acceptable part of, say, educational progress.[15]

Glover's superdrug case is a mixture of strategies, then. On the one hand, in the successive modification of the superdrug, Glover has to retain some fixed background of our values to eliminate the objections on the social-political level. Using the hypothetical case, then, is supposed to reveal our current values and commitments in our objections to the superdrug. On the other hand, part of the purpose of the example is to introduce a possibility of an alternative system of values, and discuss the general ethical issues that arise because of the powers that new technologies give us with respect to the parameters of our lives we assumed as given. So, Glover's case is simultaneously about what is doing the work here and now, what problematizes the stability of our value system by injecting the contemplation of future possibilities into our thinking, and what probes for our reactions to the pictures of such alternative futures. We are now getting closer to the idea that some thought experiments utilize fanciful imaginings in ways that are not captured in Wilkes's discussion.

[15]Think of Kramer's work as well.

4.6 Diversification of Functions

Whatever reservations one may have about Glover's methodology, the discussions of autonomy, diversity, and so on, prompted by considering thought experiments, seem to lead us to a better and more nuanced understanding of our values and commitments. Some of this may be explained by what Wilkes, Thomson and Glover agree on, namely that some thought experiments may be conclusive if we are able to manipulate the variable without disturbing the background. However, this does not apply to all uses of imagination even in Glover's own work. Often we envision possible futures because we are not at all clear what the fixed background is or should be, or are not clear how to weigh different aspects of the background against each other, and so on. If some uses of imagination help us understand *how* to fix background in the first place, or if they help us contemplate our possible futures, then we should seek other explanations of the fruitfulness of thought experiments in this respect.

Adrian Walsh suggests that many thought experiments in health care ethics can fall into the following four categories: counterexamples and reductios, intuition pumps, commitment cleavers, and "re-imaging" thought experiments (2011).

Counterexamples and reductios include thought experiments of fantastic variety that target certain universalist conceptions of ethics. According to some understanding of philosophical definitions, they must provide necessary and sufficient conditions that apply universally. If a philosopher were to claim that a particular truth holds in all logically possible worlds, he should be prepared for counterexamples based on any of those worlds. Think of the whole industry of objections and replies meant to support or disprove utilitarianism. These kinds of examples, along with what they are meant to refute, are clearly subject to Wilkes's objection.

Intuition pumps, "aim to lead us to to some general kind of conclusion from our reactions to a single thought experiment" (Walsh 2011, 179). For example, Walsh classifies James Rachel's case of Smith and Jones as intuition pump. In this case, each of these

two men walks into his bathroom intending to drown his cousin. Smith does this horrible thing, while Jones's action is preempted by the cousin's own misfortune of starting to drown. Jones does nothing to save his cousin, even though he can. On the basis of this example, Rachels concludes that there is no difference between actions (Smith) and omissions (Jones).

Commitment cleavers "are used to enhance understanding by teasing apart distinct but potentially conflated principles" (Walsh 2011, 178–179). Walsh interprets Plato's *Ring of Gyges* as a commitment cleaver to determine whether the requirement of acting justly is one of prudence or of moral obligation. In Plato's *Republic*, Glaucon wonders whether anyone will be motivated to act justly if he possessed the ring that makes one invisible. If there is a possessor of the ring who nevertheless does not commit crimes, then justice goes beyond acting out of fear.[16] Glover's superdrug, for example, clarifies which features of the superdrug we find more objectionable than others, by separating the cases in which we might have confused the objections from social quietism with the objections from autonomy.

Re-imaging thought experiments are those that "reframe or refocus" familiar debates in a new light (Walsh 2011, 179). For example, Walsh thinks that Thomson's case succeeds by shifting attention of the participants in the debate to the features of the situation that have been overlooked. Re-imaging allows us to frame the problem differently by focusing our attention on a different set of criteria by which we can interpret the situation. If we conclude, on the basis of our reaction to this case and the argumentation, that it is permissible to unhook from the violinist, and if we take the idea that this is analogous to the situation of unwanted pregnancy, we can conclude that it is permissible to abort a fetus, even if we grant the status of a person to the fetus.[17]

Of course, Walsh's taxonomy is not exhaustive; in addition, there may be cases that accomplish several of these roles at once. It would be surprising, for example, if re-framing did not also serve as a commitment cleaver. Our superdrug case can be seen

[16] Recall our discussion of this case in Chapter Two, for contrast.

[17] See Elizabeth Camp's similar view on what happens in literature in Chapter Three.

as utilizing both of these strategies, to some degree. Putting this to the side for now, however, the value of separating the functions of thought experiments may allow us to circumvent some of the popular objections to the fanciful examples in bioethics that often result in the demands of making bioethics "more empirical."[18] For example, what Walsh calls the"objection from modality" is essentially Wilkes's objection. Having provided the distinctions, we are now in a position to see that the objection applies to some categories of thought experiments, but not to others. For example, if we are after predicting the future in some fantastic case of brain manipulation, then it may well apply. But if we are after rethinking the relation between direct and indirect influences on the brain in education, it may not. As an outcome of his taxonomic discussion, Walsh urges health care specialists to continue to engage with thought experiments instead of rejecting them for their bizarreness, as long as each thought experiment is analyzed according to the goal that it pursues and the argumentative context, in which it is embedded. According to Walsh, the problem is often not in the thought experiment itself but in the failure to respect the contingent facts of the situation, in which the thought experiment is embedded.[19]

Walsh's discussion allows us to separate different functions of thought experiments that may be confused in order to avoid some of the unnecessary entanglements produced by the confusion. However, I don't think this is the end of the matter. It stops short of engaging with the dynamic connection between technological possibilities, imaginative projection and speculation, on the one hand, and the shared background of our values and practices, on the other. I discuss this connection in the next two sections.

4.7 Imagination, Practice, Background

It is not accidental that Walsh warns us about the "extraordinary contingency" of bioethical discussions. One way to understand this is to say that each case is different, and any general conclusions from thought experiments have to be treated carefully.[20] The

[18] See Arras 1991, 1994, 1997.

[19] According to him, the health care ethics is dealing with "very particular set of circumstances"(182). See Dancy 1985.

[20] See Dancy 1985 for a version of this view.

problem may seem to be with the lack of information of how to connect the fantastic with the actual, or with the wrong selection of the framing features for the comparison. While these are very important issues, this kind of presentation does not sufficiently address the role that imagination plays in these contexts in which it seems so natural and fitting.

Another way to understand the extraordinary contingency of the discussions of practical contexts has to do with thinking about the role that imagination and speculation plays in framing the background of the practical discussions in general, and in bioethics in particular. Practical contexts are both predictive and normative, present- and future-oriented, factive and fictive. For example, our policies have to be flexible enough to deal with future changes. Our understanding of the use of imagination in bioethical contexts has to reflect *this* extraordinary contingency.

Return for a moment to our starting discussion of Wilkes, according to whom our intuitions function relatively well when the background is the familiar background of the actual world. However, there is no universal agreement about what to say in many of such cases. The difficulty is that in so many cases what the facts of the matter look like and what to do about them is not settled without a decision about the background practices in which these hard cases arise. Our assessment of such cases depends, in Wilkes's words, on "careful handling and interpretation." Imagination may play a vital role in determining which parts are more flexible than others. If this is plausible, then appeals to imagination in these negotiations are to be expected in our attempts to settle which features of the background have to be fixed and which have to be ignored.[21]

So, experiments are on a continuum with the standard practice of reflecting on actual hard cases. We wonder about the background system of institutions and values that surround our decision-making. The discussions of what our future may be like, what kind of future *we* would like, etc., must be relevant to how we think of our current situation. For example, the fact that we find some future developments more desirable than others

[21]Wilkes's own proposal about resolving MPD cases is to question the assumption of the unity of consciousness—one of the criteria of personhood she starts with. But this is not the only option, and decisions about which option is more plausible to adopt may in part depend on what we think a person is, which may be tied to what kind of lives of humans we can imagine.

may play a decisive role in what kind of future is more or less likely because what we want plays an active role in how things turn out.[22] Thus, Wilkes's requirement of full relevant background specification as a condition on the fruitfulness of thought experiments cannot be sustained on the broader understanding of our practical concerns as dynamic and future-directed. In contemplating the fantastic extensions of our current practices, we are testing the ability of our practices to accommodate certain fantastic possibilities (by exhibiting where certain features of our lives are more flexible than others), and the visions of such possibilities influence what we think about what our background values are. Let me illustrate this in the next section by using Hilde Lindemann Nelson's reflections about her hydrocephalic sister Carla (2002).

4.8 Nelson's Sister Carla

The context of Nelson's discussion is the question of the boundaries of our attribution of the concept 'person' and 'personhood.' Many philosophers define personhood by giving a set of characteristics that distinguish persons from other entities. Nelson lists the following familiar examples of such features: rational awareness, self-awareness, beings subject of ascription of intentional predicates, capacity for second-order desires, capacity to construct one's narrative and so on (2002, 32–33).[23] Such definitional work personhood is usually taken to have important implications for justifying our practices with respect to persons.

But there are hard cases that make such an approach less attractive. Let's reflect on Nelson's biographical experience with her hydrocephalic sister Carla. As we learn from Nelson, Carla did not possess any of the features of personhood that often occur on philosophical lists of "conditions of personhood": she did not possess self-consciousness in any strong sense, nor did she use language, etc. Yet, Nelson claims that it was enough for the engagement of recognition by others and responses to her as a person that "she

[22]This is not to say that we determine how things will play out, of course, but to point out that we can play an active role in shaping the future.

[23]In Chapter Two, we saw Wilkes's appropriating Dennett's list of six conditions of personhood, which may be taken as an example of giving necessary and sufficient conditions for such class membership.

had experiences and sensations; she could fix her attention; and she could be comforted" (2002, 35).[24] Nelson says that these characteristics were enough for others to "hold her in personhood." Others played with her, took her on vacations, talked to her, and so on and so forth; i.e., even though Carla herself did not possess the full range of the capacities we associate with personhood, the efforts of others have involved her in the recognizably personal life: she was dressed, talked to, comforted, and cared for, as a person would be.

Nelson traces natural attitudes we have towards others to the idea of a 'form of life.' She writes: "How we think about and behave toward things of a certain type is tied to the attitude we are taught to take toward such things, and this in turn is tied to the form of life we inhabit" (2002, 33). Crucially, without the attitude we have had to learn to take towards particular bodily expressions, words, gestures, our own reactions to those expressions and what they reveal would not be what they are. Nelson writes: "If we take seriously, as I believe we must, that these states are socially mediated and that persons too are essentially social, then rather than tying personhood solely to capabilities and competences residing within the individual, we have to see it as partly also an interpersonal achievement" (2002, 34).[25] In Carla's case, what held her in personhood are the efforts of her family rather than her own contribution.

Of course, the idea of "holding some individual in personhood" is controversial. The following argument can be run as a reductio of Nelson's idea. If Carla can be "held in personhood" by the narrative woven for her by others, where do we draw the line between us, other animals, and even inanimate objects? Can't they, too, be cared for, played with, have names? Why can't we "hold them in personhood"? Without the properties that spell out the conditions of personhood, the decisions about personhood attribution start to look arbitrary. This objection becomes even more forceful if we continue the line of cases further towards the disappearance of the paradigmatic features of personhood.

[24]Notice that having experiences, fixing attention, and behaving as if comforted is exhibited by very many non-human animals.

[25]Nelson's definition: "Personhood just *is* the bodily expression of the feelings, thoughts, desires and intentions that constitute a human personality, as recognized by others, who then respond in certain ways to what they see." (2002, 34)

While Carla had experiences, could fix attention, and could be comforted, humans who enter persistent vegetative states, for example, do not. Nelson herself denies that more extreme cases like that can give us enough to hold those persons in personhood. But then doesn't it look like to distinguish between Carla and PVS patient one needs to rely on some minimal set of characteristics—precisely the idea Nelson wanted to avoid? In this case, what distinguishes the cases has to do with the capacities that Carla has, the traditionalist may point out.

Nelson's own answer to this argument is an appeal to the idea of a 'form of life'—the set of practices afforded by our mode of embodiment, history, etc., etc, into which other animals simply cannot enter (2002, 35). Nelson says that we can neither inhabit theirs, nor bring them into ours because of "the very specificity of the configuration-in-context that lets us zero in on the person's subjectivity" (2002, 34). In order to recognize the sequences of bodily movements, sensations, emotions and so on and so forth as expressive of another person's subjective states, we have to have been trained to do so. But of course, the expressions themselves are also trained responses—there is a continuum of mutual adjustments of expressions and responding to them. These very patterns are the features of our 'form of life', and they depend both on our biological and social development. Thus, the patterns of holding Carla in personhood by involving her in the family practices do not float free from the kind of face she has and the expressions it can bear, or from what is taken to be paradigmatic signs for humans for fixing their attention, for looking pacified, and so on. Such signs of personhood can only be recognized if various other conditions are fulfilled: being able to move in certain ways, to have a certain kind of face, to fix one's attention in particular ways, and so on. As Nelson explains, "the mode of being that is supported by their embodiment is simply too far removed from ours for us to draw cats into our practices of personhood" (2002, 35). I take it that this means that there are simply too many differences in basic embodiment, needs and sensibilities between us and other animals, for us to be equal participants in the same form of life.

Let's now examine Nelson's other cases on the spectrum of deviation from the normal

developmental trajectory for humans. According to Nelson, in PVS, there is no mental life we can be witness to, no signs of being comforted, no feedback that is recognizably interpretable, etc., and so nothing for a body to express, which makes our holding them in personhood inappropriate (2002, 35). Nelson's stance on fetuses is more ambiguous. On the one hand, their being propped into personhood is severely limited because "they are hidden from view"; on the other hand, the case is complicated by what she calls "anticipatory identity construction" based on the narrative a prospective mother or somebody else may start weaving about her future child (Nelson 2002, 35). Such identity construction is not yet the narrative of the individual until the baby is born, and its status is more fictional. Interestingly, though, Nelson speculates that further developments in scanning technology, which could make more *images* of fetuses available could actually push our recognitional capacities of fetuses as persons to earlier and earlier stages of pregnancy. At the end of the day, then, whether fetuses can be held in personhood may depend on our technological advances.

Now, Nelson's own position is supposed to rely on an individuals' willingness to hold another person in personhood. The PVS cases, however, present a problem for her position. It may be too quick to say that "there is nothing for the body [in PVS] to express." A living body in PVS is motionless, but so is a body in sleep, and surely there may be something for a sleeping body to express for us. At the very least, we have to acknowledge that the process of recognizing the loss of personhood is gradual. Put slightly stronger, PVS presents us with more of a candidate for recognizable human form than a fetus does. So, the difference in our reactions to PVS and to fetus cases has to do with there being little to anticipate in the PVS case. However, if we can have "anticipatory identity construction," perhaps we can also have "memory-based identity construction" that brings the life of an individual's narrative to an end. Thus, it is by no means a fact of the matter what we make of them without further reflection and interpretation.

My project does not hang on the resolution of the tensions in Nelson's view. I use Nelson's examples to illustrate the dependence of what we take facts of personhood to

be on our reactive attitudes and to show the continuity between and interdependence of reflecting on hard cases, technological possibilities, and fantastic scenarios. Starting with the actual cases of Carla, PVS, and fetus cases, we reflect on the differences and similarities that these cases have amongst themselves and as contrasted to the normal cases. As a result, we may uncover new ways of thinking about various aspects of personhood that determine our reactions in each case. In addition, by discussing the influence future technology may have on the extension of the application of the concept of personhood, Nelson has opened the door for further use of imagination in our reflection. If we are prepared to take seriously the idea that what counts as person is not simply fixed by the listing of necessary and sufficient conditions, but can be seen as an interpersonal achievement that depends on cultural practice, and if technology becomes available for radical alteration of human embodiment or psychology, then again we would have to reflect on the possibilities of incorporating this new embodiment of psychology into our form of life. As we concluded in the previous chapter, fantastic stories of disembodiment and body-switching have always been preoccupied with the possible integration of such beings into our human practices, and we can learn a lot by observing our own reactions to such attempts. The fantastic cases that philosophers use can be seen as narrowing and localizing of this procedure of using imagination to reflect on the interaction between the entity of the agent in the world, and our values. They are part of the internal critical resources of our practices.[26] Let's further develop this idea in the next section.

4.9 Back to the Future

As Buchanan et al. (2001) suggest in the opening of *From Chance to Choice*, in appealing to thought experiments, we are after two different but interrelated tasks. First, we want to assess the potential risks of this or that technology that is on our horizon. For example, suppose couples were able to design their "more ideal children" at some point in the near

[26]David Shoemaker remarked in our correspondence that the appeal to our own reactions threatens a kind of objectionable conservatism, possibly coupled with various group biases. Is such begging the question acceptable, according to my approach? I hope to alleviate this worry to some extent in what follows.

future. It would be dangerous, according to the authors, to abstain from reflecting on the ethical consequences of adopting such technology merely because at this point this technological advance is speculative.[27] Second, our imagining possible futures also prompt reflection on our current values: Do we have the ethical resources to use our genetic powers wisely and humanely? Buchanan et al. argues that what we need is a "systematic vision of the moral character of the world we hope to be moving toward" (2001, 487).

There is a clear connection between these two tasks, which brings us back to the initial justification for including a discussion of bioethics into this thesis. Risk-assessment presupposes some background of values. But since technology expands the aspects of our nature that are subject to our control, we realize that even the starting assumptions for risk-assessment are subject to change. Thus, reflecting on risks in each such case brings along reflection on the adequacy of our assessment of them. We are indeed in a peculiar situation: to assess risks we need to assume the background of our values, while the background itself is flexible because of the new technology.

This discussion echoes what we established. Some fanciful scenarios are meant to ask what would happen given this or that background and this or that technology, on the assumption of our current value system, whatever it is. But others ask a related, but different, question of how we do and would feel about the future technological developments as they change the landscape of our powers to affect different aspects of our lives, including the contemplation of the possible change in our values. Among the critical resources of reflecting on such flexible field, there are various uses of imagination that take the form of thought experiments: cases like the superdrug, the body-transfer, the human zoo, etc. Each of these cases is prompted by currently available technologies and speculatively extends the powers we have over a certain domain of human experience. Even though nobody can guarantee that these cases are going to take place in the manner described in those fanciful scenarios, or whether they would happen at all, such extensions are natural because of the practical nature of our reflection.

[27] Buchanan et al. 2001, Also See R.M. Hare 1993.

As we saw in the previous section, Nelson used actual cases to reflect on our practices of personhood ascription. In addition, their extensions to the possible technological future is almost a foregone conclusion. Consider the idea of "anticipatory identity construction" afforded by new scanning technologies that allows us to *see* how a fetus grows. Technology is changing both the intuitions we have about the "margins" of personhood, the status of the entities at the margins, and the landscape of our institutions. Given a certain fluidity of our concepts, we are bound to speculate about the extensions of our practices to the future.

While it will be said that contemplating fantastic possibilities is mere speculation, I think this does not make such speculations necessarily arbitrary. (I deal with this objection in Chapter Five.) Carefully thinking about our reactions to different degrees of modification of some parameter of our lives may help us rank the more and the less likely changes. Let's take the superdrug case. As I said, we can incorporate many of Glover's insights into our understanding of his thought experiments as testing for possible extensions of our current practices. The case is initially posed against the background of existing mood-enhancing drugs, but it takes us further to see what would happen if they were better, more widely available, had no objectionable side-effects, and so on. This is not meant to serve as an accurate prediction of what will happen or about the values we currently share. In addition to being about such current values, it is an invitation to test our reactions to a potential development in our technology. The usefulness of considering our reactions in these speculative cases comes from the fact that our reactions to the imaginary are not arbitrary: they have a natural home in dynamic practices with a certain historical trajectory and a certain degree of flexibility. While these practices are certainly flexible, in considering different spectra of our reactions to successive elimination of the objections we may currently have to mood-enhancing drugs, we put ourselves in a position to appreciate whether some objections survive this kind of reflection better than others. For example, in the superdrug case, we may have a lingering suspicion throughout the spectrum of cases that ever-powerful mood-enhancement technology will

undermine, or limit, or overpower, other ways of attaining happiness. Or we may worry about excessive fixation on pleasurableness of experience. If one of these responses elicited by the superdrug case is strong enough and is present throughout the spectrum of cases, then we may conclude that the significance we ascribe—implicitly or explicitly—to some ways of achieving happiness is an important factor we have to consider in our future policy. So, the fact that the lingering suspicion is not alleviated through the successive changes of the thought experiments in the spectrum indicates a particular constraint on how our practices are likely to develop. If another reason for resistance drops out of the consideration in some cases, there is (defeasible) reason to think that this latter feature is not as important as the earlier one. Such a procedure is not simply a form of conservatism, but also the appreciation of certain constraints in how our practices may develop. Not all such constraints are as strong as others, and such imaginative ranking can indicate to us the strength of our commitments. On the other hand, of course, pushing our reflection in this direction may show that many of our current practices should be significantly revised because they do not survive the scrutiny of imaginative reflection.[28]

Let's now look at another example—the genetic cloning of individuals for the purpose of easier indoctrination into a particular religious community.[29] The question we confront is what can be objectionable in producing individuals with desirable characteristics, some of which may make them more susceptible to particular ideas, which may, in turn, actually make them subjectively happier? (In this case, for example: the clones will have an easier time fitting the community, won't suffer from various forms of intellectual doubt, and so on. In slightly modified cases, we can think of making individuals healthier, for example.) What is the difference between the "traditional methods" of indoctrination and (on pain of incoherence) genetic means to achieve the same results? It is not, I think, accidental that this scenario is reminiscent of Aldous Huxley's *The Brave New World*,

[28] An illuminating discussion of this theme is found in Cavell's discussion of slavery (Cavell 1979, 376).

[29] Let's put aside the objection that genetic determinism presupposed by this thought experiment is incoherent (Sober 2001). We can be flexible with the content of the proposed purpose of the modification, if we need to: it is not incoherent to think that *some* of the reasons people may want to produce clones are more feasible than others.

in which reproductive technology allows people to create individuals with characteristics desirable to the well-ordered society. I take it that many of us will be against such abuses of technology because whatever the freedom the religious organizations may enjoy, the freedom of an autonomous individual to make her own choices is not something we want to give up even at the cost of her having an easier life, even though the visions of ideal social order are very tempting. This is surely no news, but Huxley's elaboration of what the world may be like without it may suggest the strength of the grip that the idea of individual autonomy may have on us. So, again, our speculations about possible genetic modifications may show that some possible genetic modifications seems more acceptable than others, depending on what we think is deeply important about uniqueness and autonomy of individuals. Again, we certainly do not know how the cloned individuals will view the situation. Part of the power of Huxley's example, however, comes from the fact that we suspect we actually know that many of the clones would feel as content as they could be—certain ideas simply would not cross their minds—and *that* vision moves us. The fact that there is a broad agreement that this loss of critical capacity is not something desirable shows, again, a certain accepted view that some things we may not be willing to give up even for a very well-ordered social organism.

It may be asked at this point, whether what we are supposedly learning here could have been obtained with more reliable resources. We could have looked at how people think about things like autonomy and social good now, without appealing to any fancy imaginations. What does the fantastic add? I have already mentioned that the fantastic elements may give us resources to rank the importance of different clusters of values and ideals to us. For example, we learn that autonomy is the structural feature of many of our other evaluative practices, and our concerned reaction to the proposals of the bright futures when it is eliminated for the sake of other things shows us more than the actual cases can.

In addition, we should question the assumption that what we currently think is somehow independent of our imaginative flights. Our current values, ideals, and self-

understanding have not been formed in isolation from thinking and speculating about the future. Endorsing this idea, we open up to the view that some projects of self-examination have to do with imaginative projection into the future possibilities, and our ability to participate in this kind of cognitive engagement may not be reducible to the induction from the familiar actual cases. In part, the previous chapter was an exploration of the imagination-driven resources of self-understanding of our culture. The fantastic is one of the tools for self-examination. As I said at the end of the previous chapter, there is a mutual dependence of knowing how to go on in applying our concepts to the fantastic cases and actually engaging with fictional stories of fantastic transformations. The act of thinking through the fantastic cases does not simply uncover knowledge that we possess prior to the act of imagination. Instead, we can think of the imaginative engagement as actively shaping what we learn. [30]

4.10 Conclusion

At the outset of this chapter, I gave four reasons to justify the inclusion of this discussion into the thesis. Bioethical context of thought experiments was supposed to narrow the question of interest to the interaction between the entity and the practical world; this narrowing was in turn connected to the notion of guided and constrained speculation; in addition, it was supposed to bring to the front of the discussion the idea that our practices are both historically and culturally conditioned and have a certain built-in flexibility; and finally, the discussion was supposed to have instrumental value of bringing the fantastic closer to home because bioethics is located at the intersection of the past and the future—the idea we may have missed in limiting our discussion to the literary fictions.

The first two reasons are the most important, and I hope that my discussion has given a sketch of how to make good on the promise. As I said, bioethics is at the intersection of technologically-driven changes of the boundaries of human entity and our values, and there is dynamic interaction between the two. Since literary fictions are not embedded in

[30] More needs to be said about this, but I hope to say more about this in the next chapter when I come to the discussion of thought experiments in personal identity.

the context of the scientific investigation, bioethical discussions focus on more narrowly selected localized clusters of concerns that are particular to the biomedical discipline. I hope the examples given sketch the idea of guided and constrained use of imagination that complements the picture of the cognitive significance of fiction from the previous chapter.

However, the more difficult task is to establish that speculative engagement with bioethical scenarios—however narrowed and limited by the context of their occurrence within an institutional setting—can result in speculations that are not merely arbitrary. Starting with the discussion of Nelson, and following up on it in the previous section, I gave some examples of the procedure of ranking different clusters of responses to the fantastic cases. As I said, such ranking, although fallible, can produce preliminary lists of features that are mutually dependent, and the presence of which throughout the range of cases can indicate the degree of their importance. In the next chapter, this feature will be explored in more detail when we apply the lessons from the previous two chapters to metaphysics of personal identity.

Chapter 5

Fictioning Thought Experiments

5.1 Introduction

Let's briefly review the dialectic of this thesis up to this point. Thought experiments are accepted methodology in the philosophy of personal identity. But there seems to be a tension at the heart of the method: roughly, how can thinking about what is not the case give us knowledge of what we most fundamentally are? Even though thought experiments in personal identity cannot fit the scientific model, as we saw in Chapter Two, I argued that their cognitive significance is not tied to it. In Chapters Three and Four, I considered two contexts in which the fantastic scenarios are cognitively significant despite not fitting the scientific model of success for thought experiments. By putting the fantastic "variable" in dynamic interaction with the background world, literary fictions and some uses of imagination in bioethics bring to the surface the structural conditions that make lives possible and sustain them, thus showing the possible configurations of the changes that the clusters of such features can sustain. The possible predictions of what changes these features can undergo are defeasible—we are not making secure predictions here. However, I would argue that as long as they withstand informed criticism, their discussion does not require any better grounding.

In this chapter, I will take these lessons back to personal identity. The insights from the previous chapters will have to be applied with some modification. As I have been

saying, the idea is not to replace philosophical scenarios with literary fictions. Since my proposal is methodological, I suggest that we look for the cognitive value of philosophical thought experiments in the kinds of details we have seen at work in literary fictions and in bioethics. This changes the question that we put to thought experiments. Instead of directly asking about the persistence conditions of this or that entity, we are interested in the more general question of whether the possible world presented to us in the philosophical scenarios is coherent. Addressing it requires further development of the thought-experimental background. In some cases, such "reconstruction" quickly becomes problematic, signaling that different structural aspects that we associate with the continuation of a given person's life come into conflict with each other. The conflict shows that we may not be able to project the kind of general understanding of persons that we have into the possible world under consideration. In other cases, the task of the background construction of another possible transformation does not lead to incoherence, but points to further questions we may want to resolve before proceeding. Both types of results are important because they reveal—during the discussion of the practical background of a given possible world—which clusters of features associated with personhood can be tweaked, mixed and matched, and so on, showing us important dependencies between aspects of our being. I will call this approach yo background construction the "literary model of thought experiments."[1]

Here is how the chapter unfolds. After briefly describing some general features of the model we have seen at work in literary fictions and in bioethics in section 2, I concretize my general observations about the model. In section 3, I spell out Parfit's division case. In section 4, I put the literary model to work by suggesting a possible narrative background development of the thought-experimental background, which I believe compromises Parfit's conclusion and suggests a different outcome. In section 5, I sketch the implications of a similarly detailed discussion of teletransportation to show the flexibility of the model. In section 6, I discuss some of the costs of adopting my model, especially

[1]No heavy water should be made of 'model' here. I don't mean model in the scientific case.

the threat of relativism one might associate with it. Without resolving the issue, I give reasons to think that the threat is not fatal.

In Chapters Six and Seven, I address two theoretical objections to the model: the idea that by bringing practical details to bear on metaphysical investigations, I have changed the subject, and that I have been assuming that there is some unified metaphysical relation that supports all of our practical concerns.

5.2 Thought Experiments and the Background

Our predicament as living beings is change. Some changes are radical enough to make us wonder whether the person can survive undergoing them—consider the full-blown loss of episodic memory or entering a persistent vegetative state (PVS), for example. Typically, thought experiments are used as tools of controlled experimentation to determine more precisely what aspects of our lives can be responsible for our judgments whether a person survived a given change: we suspend one or another feature of our lives, while keeping the rest of the picture mostly intact. Based on our intuitive reactions to the case, we draw conclusions about the corresponding psychological or biological features that presumably secure this judgment.

To illustrate, consider something like a schema for the traditional approach.

1. We assume that judgments of personal identity are intimately connected to judgments about the holding of practical concerns. For example, in Locke's classic case of the Prince and the Cobbler, as it is typically interpreted, the judgment is that the person of the Prince goes wherever goes the inheritor of his concerns, such as responsibility, compensation and so on. To generalize, our judgment that some person is responsible for my past actions, should be compensated for them, etc., is taken to imply that this person is identical to me (barring some exceptions).[2]

2. We now imaginatively suspend one or another feature of our life. Typically, some crude distinction is made between the psychological and the biological features. For

[2]In the following chapters, I discuss various criticisms of this central assumption.

example, some artificial means of separating embodiment from psychological relations is invented: xerox-machine, psychological-state recorder and replicator, memory-wiper, and so on.

3. We ask: what is the intuition about the outcome of the case? We will think that (a) in some cases the person survives the ordeal, or that (b) in some case the person does not survive the ordeal.

4. We conclude that: (a) because of (1), the outcome preserves my person, and whatever happened is as good as survival; or (b) because of (1), the outcome does not preserve my person, and the output of the device is as bad as death.

Of course, the subsequent discussion is much more detailed, but the sketch affords, I believe, a view that the intuitive reactions about the holding of our practical concerns play the role of evidence for the preservation or cessation of identity. That is, judgments about practical concerns are directly tied to judgments about identity.

Notice that there is a host of assumptions behind this procedure. The most general is that our intuitive reactions behave in the manner of generalizations based on evidence. In addition, there are experimental-design problems having to do with the difficulties of setting up the controls for self-reporting due to various priming effects, the differences in respondents' experiences, and so on, which are very difficult to rule out as swaying the judgments this or that way. Furthermore, there is an assumption here that our intuitive reactions about fantastic cases are transferable and applicable to what we think of ourselves. One may think that such problems doom any systematic and fruitful investigation.

Investigating these problems is not my project here, however, even though tracing the possible points of convergence between this thesis and other such investigations would be a fruitful addition to the personal identity discussion. Bypassing such issues, I focus on the general shape of the procedure, which is reflected in the general question we ask when we perform thought experiments. According to my proposal, the question is not the direct question whether a person survives teletransportation, with the subsequent discussion of what the judgment implies—for example, that all we need for survival is psychological

rather than biological continuity since the latter is broken in teletransportation. Looking at literary fictions and bioethics has made it clear that discussions of individual aspects of our lives are always tightly intertwined with various other features and the general contexts of our form of life. Thus, our inquiry about survival has to presuppose the conceptual background in which questions of survival can even arise. I suggest that we ask whether the imagined case is coherent and whether we can truly envision the full-blown scenario of the vicissitude. To do this we will have to inquire what further possible-world background we need to investigate in order to discuss this case and to investigate the interactions between different clusters of our concerns that we can observe in such cases. This set of questions cannot be answered without imagining the general social and practical background of the thought experiments. As a payoff, such details uncover resources for understanding the interactions between different clusters of concerns and practices in which we are embedded as living things. Having a better understanding of these aspects of the discussion of personal identity can put us in a better position to understand both the traditional metaphysical questions and a non-traditional role that thought experiments can play in addressing them.

One might think that such a move is problematic on the following grounds. We started with the problem that our intuitive reactions to fantastic cases are subject to bias, are easily manipulated, and may in general not track anything related to personal identity.[3] In light of this, what I suggest may seem like it can only worsen the problem by adding even more problem-generating details. I realize the difficulty, but submit that this point does not favor the standard procedure, but perhaps reminds us that there is an element of speculation whenever we are talking about imagining hypothetical situations. All such methods are in the same boat. In order to claim that my model is worse off, the objector would have to show that the difficulties that come up in further elaboration of thought-experimental backgrounds are somehow multiplied. But this can only be done by analyzing each individual description of the cases generated by the model that the general

[3]These objections can be brought together under two labels: the cognitive diversity problem (Stich 1988) and the calibration problem (Cummings 1998). Also see the discussion in Baz 2012, Chapter 3.

objection does not capture. I suggest we look at the results of the procedure before giving up on the project. In addition, I want to challenge the assumption that since it is difficult to disentangle different strands or aspects of our lives, the bundle inherits the sum total of the difficulties that arise with respect to each individual feature. On the contrary, I think the difficulties may be the result of assuming that the philosophical analysis of the notion of survival must provide a set of necessary and sufficient conditions for life/death and must apply with full generality (See Chiong 2005). This picture of philosophical analysis is not the only option, and so the objection stemming from it may not apply to views that do not start with such a picture. This does not eliminate all the difficulties with the view, of course, but it does postpone the question at this stage in the dialectic.

In the following sections, I discuss Parfit's famous cases of fission and teletransportation to substantiate these abstract remarks.

5.3 Fission

5.3.1 Parfit's Discussion of the Case of *My Division*

In Parfit's version of this standard case, we are asked to imagine a terrible accident, as a result of which my two brain hemispheres are successfully transplanted into the bodies of two of my brothers, who look very much like me. As an outcome, "[e]ach of the resulting people believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me" (Parfit 1984, 254). Parfit asks what happens to him in the story.

Intuitively, there seem to be four options on the table. I survive as one, another, both, or neither, of two descendants. First, suppose I am one, but not the other, of the two survivors. This cannot be right, however, because we assumed that each is equally psychologically continuous with me: so there is no justifiable basis for the preference among the two. This takes care of the first two options. Second, can I survive as both? But how could two people comprise the third? Imagine that these two people fight a duel. If one dies, is it murder or suicide? This option seems to twist our concepts too

much, according to Parfit (1984, 256-257). Lastly, should we say that I do not survive? Consider actual cases, in which hemispherectomy—removing one brain hemisphere from a patient's skull—is performed in hospitals. Supposing that the person's functioning is restored to some extent, it seems right to say that that person survives. So, a person can survive with one hemisphere. Parfit argued earlier that he would survive if his brain were successfully transplanted.[4] So, the combination of the two allows him to say that one can survive a single hemisphere-transplantation. But now add to this the idea that the hemisphere that is usually discarded can be successfully transplanted. Supposing so, Parfit asks: how could doubling the relation in which I stand to the single survivor be a failure to survive if doubling does not change the intrinsic nature of the relation between the original and either of the resulting people (1984, 256)?

In *Reasons and Persons*, Parfit argues that it is an empty question what distinguishes the four possibilities: each of the four is an alternative description of the same outcome, about which we know all there is to know: that there are two survivors who are psychologically continuous with the original (1984, 260). He does think that we *choose* that the best description of the outcome is that neither of the survivors is the original (Parfit 1984, 260). In a later paper, Parfit argues that *I* won't survive this kind of ordeal. There, he says that it is false of each survivor that he is me because the relation of identity is one-one: two people cannot be identical to each other (1995, 42). Moreover, "When this relation holds between me now and *two* future people, I cannot be called one and the same as each of these people" (Parfit 1995, 43).

Since according to him the original question whether I survive division is empty, Parfit suggests we ask a different question: "What ought to matter to me? How ought I to regard the prospect of division? Should I regard it as like death, or as like survival?" (1984, 260) Parfit argues that it is irrational to regard the prospects of division as like death.

[4] "Just as my brain could be extracted, and kept alive by a connection with an artificial heart-lung machine, it could be kept alive by a connection with the heart and lungs in my twin's body... If all of my brain continues to both to exist and to be the brain of one living person, who is psychologically continuous with me, I continue to exist. This is true whatever happens to the rest of my body" (Parfit 1984, 255).

As stipulated, each of the fission-products will have the full psychological profile of the original, and one hemisphere transplant surely secures survival. But then, as Parfit claims, "[i]t cannot be the *nature* of my relation to each of the resulting people that, in this case, causes it to fail to be survival. Nothing is *missing* " (Parfit 1984, 261). But if we agree on that, how could doubling that relation result in death? And, if it is not death, then it is survival.

The shocking conclusion follows. Division seems to preserve what matters to us in our survival: that our psychological life, with its projects, anticipations, memories, and so on, continues. However, division cannot preserve identity because if I am identical to both of my survivors, by transitivity they are identical to each other, which is impossible. But if they are not identical to each other, neither am I identical to either one of them.[5] So, "identity is not what matters in survival" (Parfit 1984, 264).

Of course, Parfit acknowledges that having a duplicate will be "uncanny", and will raise practical problems, causing what Parfit thinks are "minor disagreements" (1984, 264). For example, it would be "pointless" for both survivors to write the same book; the woman who loves Parfit "could not give to both the undivided attention that we now give to each other"; and so on (1984, 264). On the other hand, Parfit thinks that the survivors could effectively pursue conflicting interests, or resolve the problems caused by the incompatible ambitions of the original (1984, 264–265). Whatever the complications, however, according to Parfit, these practical matters cannot change the rationality of our conclusion that identity is not what matters in survival.[6] Whatever we think about dividing brains, the question Parfit asks is this: what ought *to matter* to me in my division? I will question the legitimacy of Parfit's answer to this question. As we will see, what Parfit calls minor disagreements—about the consequences of fission for the lives

[5] If A=B, A=C, then B=C. But B=/C. So, A=/B or A=/C, or both.

[6] In Chapter Two, I discussed various problematic assumptions that Parfit makes: the assumption that the functions of the two hemispheres are identical (i.e., the full redundancy of the hemispheres), that the brain stem and the periphery do not play an important role in psychological functioning, and so on. Wilkes makes a persuasive case that, given what we know about biology and the functioning of human brains, the situations depicted in thought experiments like brain bisection are more than just technologically impossible. We agreed to grant these assumptions and pursue a different strategy.

of fission products—reveal a questionable assumption about the nature of the question of what matters in survival and thus changes the parameters by which we can evaluate answers to it.

5.3.2 Thinking About Lives in the World of Fission

Parfit suggests that to answer the question of what matters in survival it is sufficient to look at the intrinsic relations between me and the fission products. Let's review his argumentation. If there were only one future continuer standing in this relation, I would regard it as survival. But the *nature* of my relation to each of the two continuers is exactly the same as the nature of my relation to my future self in ordinary survival. If one were to argue that division is like death, one can only fault duplication itself. That is, one has to show that the fact of duplication changes the nature of the relation itself. That is indefensible, according to Parfit:"[t]he only difference in the case of division [as opposed to the ordinary survival] is that the extra years are to run concurrently... [I]t cannot mean that there are no years to run. Double survival is not the same as ordinary survival. But this does not make it death. It is further away from death than ordinary survival" (Parfit 1984, 262).

According to Parfit, what matters is relation R: "R is psychological connectedness and/or psychological continuity, with the right kind of cause" (1984, 262). Why do we have trouble accepting this view? The reason we are tempted into thinking that duplication is a problem for survival, according to Parfit, is our pre-philosophical commitment to the idea that there is some "deep further fact" of identity, in addition to the relations of psychological connectedness and/or continuity (Parfit 1984, 262). But since there is no such deep further fact of identity—or so argues the reductionist[7]—all that matters is there as long as relation R is there. Identity is just R holding uniquely. So double survival is further away from death than ordinary survival because the number of experiential connections that ordinarily secure survival is doubled.

[7]According to reductionism, there are no facts of personal identity over and above the facts about psychological and physical states.

Suppose, however, that you think that it is precisely uniqueness that makes identity (as R holding uniquely) matter. Parfit's response is this: "[i]f I will be R-related to some future person, the presence or absence of U[niqueness] makes no difference to the intrinsic nature of my relation to this person. And what matters most must be the intrinsic nature of this relation" (Parfit 1984, 263).

I will not argue against reductionism. But wherever one stands with respect to that issue, Parfit's argument is question-begging. Our question was what ought to matter in survival. Whether there is a deep further fact or not, Parfit has not argued that the only basis for mattering is tied to the intrinsic nature of one's relation to the continuer. As we see from the quote above, Parfit simply asserts that uniqueness does not matter much, which is not convincing to somebody who holds that it does.[8]

It is important to note that Parfit, at least by 1995, changes the argument slightly so that the objection above may not arise.[9] As before, the doubling of survivors in the fission case does not change the nature (or the content) of the relation, in which the original stands to each of the survivors. So, in each of the two cases, the relation contains what matters (Parfit 1995, 42–43). Parfit is clear that identity fails in the double case, but he also adds that "even if we won't survive, we could have what matters *in* survival" (1995, 44, fn44). This restatement seems to admit that an objector may resist calling what the content of relation R secures survival; i.e., Parfit admits that an objector may deny that this is survival. But this denial has nothing to do with the actual facts of the situation, according to Parfit; the difference has to do with our conceptual scheme (1995, 45). Certain concepts with which we operate and the connections between them—such as the idea that our survival requires identity—tend to mask what Parfit calls reality, which

[8]This line of thought is understandable if we think how we got to this point in the argument. Recall that the double case follows Parfit's discussion of the single cases of brain transplantation and one hemisphere removal. We judge that removing one hemisphere secures survival. But what is the justification for this judgment? According to Parfit, it is the relation between the survivor and the original: the relation R. Thus, Parfit links the judgment of survival solely with the preservation of the relation R, and the subsequent discussion is about the nature of that relation. We are told that doubling cannot change the *nature* of this relation R to the two continuers, because the nature of the relationship is the same, just doubled. And since the judgment of survival is now linked to the nature of the relation R, which is preserved in the double case, whoever denies survival in division seems irrational.

[9]Thanks to David Shoemaker for pointing this out to me.

is, according to him, captured in reductionism. But *that* difference, according to Parfit, is trivial. The loss—of talk about identity, or even survival—is not important because what matters in survival is preserved twice; there is nothing that will be missing in terms of the relations that are important to us anyway, now and in hypothetical cases, when we will not be able to describe the situation as that in which *I* survive (Parfit 1995).

However, notice that the argument that what matters in survival is preserved in division only works by assuming that it is only the intrinsic relation that matters, and so it begs the question against the objector. If we thought that relation R is not the only thing to consider in the case (if we thought, for example, that factors external to the intrinsic relation may go into our judgment of what ought to matter), we would not be convinced by Parfit's argument (see Gendler 2004).

How can one go about questioning whether Parfit has identified all the elements of what matters in survival? Well, we have to try to address the question of what matters by asking for whom it matters. The basis of our answer to this question will have to do with thinking about the more general background that surrounds judgments of survival, including what can be thought of as the consequences of fission for a human life.[10] I don't mean to draw any connection to utilitarianism here. Rather, we should think of the exercise as asking for a more completed picture of the world in which the transformations occur, including their dynamic development.[11]

Let's think more carefully about the single case. I think if we probe the link between the judgment of survival and the preservation of relation R, we can put ourselves in a better position to understand our resistance to call division as good as ordinary survival on either version of the argument. Consider a single hemisphere removal. It is true that we can remove a hemisphere and have a functioning individual return to the social

[10]My account closely resembles Susan Wolf's in "Self-Interest and Interest in Selves." Wolf argues that the grounds for settling the question of whether ordinary survival is "about as bad as division" or that division is "about as good as" ordinary survival are not metaphysical and instead have to do with the surface of the world: roughly, the evaluative practices particular to our form of life. As we go on, the differences between our positions will become clear. I discuss Wolf's position in the next section.

[11]What else can we do, really, unless Parfit's question "what matters" is idiosyncratic and is not addressed to us all? (It might be that Parfit speaks from a position of an ideally rational observer, whatever that might come to, but at this point in the argument it is not presented like that.)

world: the operation is like amputation. If the consequences for the brain function are minor due to the full redundancy, the individual will seamlessly integrate. This is surely unobjectionable. But does it follow from this that what matters here—for our judgment that the individual survives this—is solely the *nature* of the relationship of the individual to his continuer, namely R? Why is the judgment in this case unproblematic? Well, it depends on a very large number of background conditions that we assume without reflection and on the expectations we have about the outcome of this case. For example, we assume that the removed hemisphere does not have some independent status, that its value is contingent on the role that it plays in the life of the survivor and its value. These assumptions, among many others, form the background of the intelligibility of judgments of survival in the first place. Since this shared background is implicit, our discussions tend to focus on very specific details, while the contribution of the background then may drop out of consideration.

I think one way to understand our resistance to call division as good as ordinary survival is that our broader conceptual understanding is thrown off by the case. Our resources for answering the question of mattering can be found in the further development of the narrative of the products of fission that we have seen at work generating knowledge in literary fictions and bioethics. At this point, let's put this approach to work to see whether we can get some insights by further story-telling—the insights that Parfit might have missed by not taking them too seriously as "minor disagreements."

I think at least some readers of Parfit are struck by the poverty of his observations about the *impact* of fission on the products' relation to the original. Even though we ourselves certainly cannot predict how the lives in the possible world of fission will in fact go, we can use the development of further background details to problematize Parfit's assumption that what matters most must be the intrinsic relation to the future person. Let's start by thinking more seriously about our relations with others based on what Parfit himself says. Parfit assumes that the woman who loves him will love two Parfits, but won't be able to give each her undivided attention (1984, 264). I think it is actually much worse

than Parfit admits here. First, if the notion of love is tied to absolute uniqueness of the loved one, we may wonder if fission would have a destructive effect on the notion of love, compromising Parfit's use of it in his discussion. But suppose you think this is overly romantic. Then consider marriage, children, inheritance, bank accounts, among other civil and financial accompaniments to love, in the full glory of the detailed description. For example, suppose your husband undergoes fission. When both of his descendants come home to "their" wife and children, what exactly happens to each of them and the family? Do they both now the family dinner, pass the salt, take the kids to school, etc.? Or do they divide their duties? How could they? And to whom should the kids turn for advice? Both? How would they keep track of each father when they have to report back on how they have followed the advice? And what about the wife?[12]

With respect to some of the resources of life, division is possible between two equally qualified competitors. We could divide money and some other possessions that are similarly quantifiable. But could we divide long-term projects, like dissertations, books, pieces of music, bringing up children, or building a life together?[13] It is more plausible to think that with respect to such "items", we are confused when we try to frame them as divisible. Notice, though, that most of the stuff of life that matters is in fact like that and not like the divisible material possessions.

It comes as no surprise, then, that in fictional narratives that attempt to come up with some coherent story about the doubling of this kind, one of the products often heroically expunges himself from any connection to the business of living the same life. (Or he dies.)[14] This brute force solution is, we must admit, one way to go about this situation. But, as Wolf points out, there are reasons to regret such a move. One of the individuals will probably suffer the severe trauma of being *completely severed* from his previous life, and it is not clear whether most of us would find the psychological strength to keep our distance. We should also not forget about the damage to the one who stays behind to

[12]For further discussion of such consequences, see Wolf 1986.
[13]For further questions and more details, see Wolf 1986.
[14]Consider the recent movies *The Sixth Day* and *Moon*, for example.

pick up where the original left off. His knowledge of there being an outcast may in itself be rather disturbing and have dramatic consequences (Wolf 1986, 716). At any rate, it is an open question what psychological changes follow the realization that there is another competitor for your resources out there.

In fact, Parfit recognizes the problem, if not its magnitude. For example, in a separate section titled *"Am I a Token or a Type?"*, he admits that much of what we value would be under threat in a world where fission occurred (Parfit 1984, 294). However, he still argues that what matters in love, *as long as we make sure that there is only one Replica at a time*, is the continuation of the chains of psychological relations that secure memory, anticipation, and so on, whatever happens to identity (Parfit 1984, 297). I think this remark shows both the importance of uniqueness and Parfit's misunderstanding of what is important about it as a structural feature of our whole way of life.

Recall (from the Introduction to this thesis) our starting point: the very intuitive connection between our practical concerns and personal identity. Thinking of ourselves as moral agents who must live among others with particular historical trajectories that are recognizable and identifiable over time presupposes a stable unit around which these concerns center. My questions about the dividing husband and father were intended to reveal the social dimensions of the crisis when this presupposition is threatened by fission.

To further show the repercussions of fission, let's think of the connection between these reflections and the Extreme Claim, the idea that psychological continuity on its own cannot serve as grounding for our practical concerns. Why should I care that there will be somebody who is exactly like me, but not identical to me, who will finish this project? (Parfit 1984, Schechtman 1996, Shoemaker 2007). Why should R matter, if it is replicable? Of course, we can add the non-branching clause to the psychological criterion of identity, as Parfit does, but that move concedes the point. Again, it is important to ask what the very natural idea of mattering to or caring about presupposes. Caring for oneself or another person, for example, presupposes stable practices of reidentification. In caring, among many other things, we have to imagine and anticipate our future, align it with

that of others. If splitting became widespread or happened randomly, I think we would not be able to apply our current notion of mattering or caring to these scenarios as easily as Parfit says. How could I choose between promoting this or that future when I could have different futures, and have them several times (as additional thought experiments can surely suggest)? Should I care about this mother or that one? Can I care for both as mothers? What kind of intentions would I have to form in order to build a life with such intentions?

At the end of the day, I think we fail to get very far in answering these questions before the story in which such answers can be given breaks down or loses its interest as a story about *us*. This failure shows that the consequences of the fantastic transformations that Parfit envisions strike at the core of our self-understanding. Attempts to tell the stories of fission survivors—and the difficulties that they will prompt—will bring to the surface the constraints on the intelligibility of our surviving fission.

In sum, discussing the impact of fission on the lives of others gives us one way to criticize Parfit for not seriously reflecting on the differences between something's being a "practical problem" and something's being a fundamental presupposition for there being such practices as ours in the first place. Fission results in there being more than one referent in statements of our practical concerns, and it is rather an understatement to call this a "practical problem." If we take seriously the task of filling out what impact this change will have on our practices, we discover that it undermines the very idea of there being particular relationships between unique individuals. Thus, duplication seems to require a completely different set of practices and concerns.[15] Contrary to Parfit's assertion that "nothing is missing", everything that constitutes the evaluative practices that we need to address the question of mattering seems to be missing. As I have been arguing, the mistake is to think that the notions (vocabulary) of compensation and anticipation, of projects and cares—and in general the vocabulary of mattering—can be kept, while we drop the idea of there being a deep difference between persons. Caring,

[15]Parfit is aware of this. His project does have radical implications for our deeply held assumptions about the fundamentals of ethics.

responsibility, and so on, as we understand them, are not simply additional relations that we can super-add to the basic "personhood-relations." Rather, the intelligibility of the notions of personhood and practical concerns are holistically intertwined.[16]

I hope that this discussion shows that the entire system of beliefs and practices built around the notion of a single individual as the unit of our practical concern is compromised by the implications of fission from the point of view of an individual's life. Moreover, these details undermine Parfit's assurances that there will be no difference, from the first-personal perspective, between ordinary survival and fission-survival. Once we start telling the story, some differences of the kind I have been describing will show up, and these differences will affect the experiences of each survivor. For example, we can ask what each survivor's stance towards the veracity of their own memories will be, since there will always be another person with the same memory. So, developing a story that is responsive and responsible to these concerns, a task glossed over in Parfit's work, reveals serious and possibly intractable problems.

In sum, a more careful spelling out of the consequences of fission shows problems with the idea that division is as good as ordinary survival. The possible solutions we saw to these problems are to either cast off one of the survivors or divide the resources of one's life between the two of them. But each option is problematic and fails to address the objection. Parfit argued that the prospects of fission are as good as those I have in ordinary survival. My analysis shows that this conclusion is unwarranted.

5.4 Teletransportation

I claimed that developing the background exactly by trying to tell the stories of thought-experimental survivors shows the outlines of the conditions necessary for our continuing to apply our conceptual scheme to something we recognize as a familiar form of life. The form that thought experiments already take—of incomplete stories and partial sketches

[16] We can concede that all problems of living are in some sense "practical problems", but some of them arise within practices, while others require complete overhaul of the old ways, including the conceptual apparatus used to describe the old practices. Fission is of the latter kind.

(that can be further specified and elaborated)—indicates our interest in manipulating not simply the abstract relations of psychological continuity as such, but their concrete exemplification in a given life. The concrete details of the background world thus give us an idea of how our concepts would play out in concrete situations. This can be seen as an exercise in exploring the limits of our concepts, finding where they may break at the limit of intelligibility. In the case of fission, the fact that we cannot imagine how to construct the post-fission world as recognizable for persons like us shows that whatever those persons may be, they are fundamentally not persons like us.

As I said earlier, the difference between this approach and the standard approaches to thought experiments can be seen in the question we are asking in each case. According to the model I am suggesting, we are not after judgments of personal survival or continuation based on the presence or absence of one or another individual aspect of our lives. (According to such a view, our intuitive reaction that responsibility and compensation concerns can follow psychological continuity in cases where our same body is not preserved signifies that we are fundamentally our bodies.) Instead, the literary model of thought experiments asks the general question of the coherence of vision of our form of life in the world in which we posit the fantastic vicissitude. The question—is the imagined scenario presenting us with a coherent vision of a form of life we would consider to be lived by persons like us?—calls for a different protocol for a discussion of such cases. This means that if my discussion in the previous section is found lacking in some respects, this won't undermine the model itself. Convinced by such an alternative vision, we may have to refine the conclusions we achieved, but the model used in both cases is still the same and will remain standing.

To further develop and clarify this point, let me consider what I take to be a somewhat harder case of teletransportation, again borrowed from Parfit. As we will see, the case does not seem to deliver as conclusive a result as fission.

Imagine that you can travel to Mars by being teletransported. The scanner will destroy your brain and body, while recording the exact states of all of your cells, which

will be transmitted at the speed of light to Mars, where a brain and a body just like yours will be created out of new matter. Your wife reminds you that she has been teletransported to alleviate your worries about your prospects. After the procedure, someone (psychologically just like you) wakes up in a new body after a momentary loss of consciousness, just as expected (Parfit 1984, 199–200). (Teletransportation has a *Branch-Line* version. Several years later, you go through the process again, but this time you do not lose consciousness when you press the button. The new technology does not destroy your brain and body, but still teletransports the scanned information to Mars. You can talk to this person on Mars: he will look like you and have the same psychology. Here on Earth you will soon die, but the person on Mars will be healthy. I won't go into details about the Branch-Line version, but it should be clear that again the analysis will proceed by speculating about the background world in which such a vicissitude occurs.)

Parfit claims that, from the first-person perspective, teletransportation-survival is not different than any normal survival situation: both in terms of the content of our experiences, and socially, nothing (or very little) will be missing after the procedure (1984, 224, 242). In terms of awareness, the replica will possess the same kind of access to his psychological states as the original does, and the content of the replica's memories, intentions, and so on, is not going to be different from the original's. The assumption here seems to be that since the singleness of the life's trajectory is preserved, then the social implications of teletransportation should be minimal, according to Parfit.

How can we continue telling the story of a teletransportation survivor? And does it run into the same kind of incoherence that plagued the fission case? I am not so sure that it does. I will not test the reader's patience here by actually telling the whole story. I do trust that any such story that we could tell would converge on resolving the issues of the status of the normal cause of our continuation and the notion of mortality—the two ingredients that seem central in such cases. Instead of telling the story, let me then directly take up each of these ingredients and explain why I think the answer to our main question of coherence is inconclusive. Supposing the operation of teletransportation

a seamless transition, we assume that the difference will not lie in how things seem to the survivor, but in how we envision the implications of the difference between ordinary survival and replication playing out in our practice.[17]

To try to push Parfit on two of the points just mentioned, let's ask a different question that will eventually lead us back to the main issues. Can we maintain our current distinction between massive delusion and legitimate memory continuation in the face of teletransportation? Or: if ordinary survival is as good as replication, as Parfit says, is perfect delusion almost as good as legitimate memory (if we could make the body look the same)? A different way to put it is this: what is the source of our thinking that the protagonist's worries that *he* won't survive the ordeal are coherent? Compare teletransportation to regular waking up after a night of dreamless sleep. The degree of our conviction that in sleep we are not replaced by a perfect replica is not based on any proof or evidence. Instead, it is a basic given presupposition of our lives. (It is no use to point to some scientific evidence of continuation of processes in the body as serving as a justification for it—we would first have to assume that these factors count as evidence for what we are trying to prove. So, justifications via the evidence from neural processes come too late: our practices presuppose that we continue in the morning. The idea that a person does not go out of existence in sleep is at the bedrock of our understanding of our lives.)

That teletransportation is worrisome or even terrifying seems to be based on the idea that the destruction of my brain and body amounts to death because they destroy the normal cause of our continuation. But we already know that the bodily continuity is a matter of continual appropriation and renewing of organic matter. And we know that various parts of the body can be replaced by new parts. Suppose further that technology changes enough to the point that we have much more control over these elements. Phrasing the question in Parfit's terms, then, what kinds of continuity would matter for the judgment that one could survive teletransportation? I think that if the degree of our

[17]Again, though, it may depend on the surrounding circumstances. Suppose you found out you were replicated. Your reaction to this news is not, I claim, independent of the cultural background in which you are embedded.

trust in replication as the process of complete overhaul were consummate with our trust that we are the same in waking up each morning, then we could think replication is a kind of survival.[18] Importantly, this seems plausible as long as we do not introduce a social competitor for the place that a given person occupies in our single-survivor-world (like on the *Branch-Line*). I take it that mentioning the wife in the scenario—who assures our protagonist that nothing is wrong with her despite her having undergone multiple replication—is meant to point to the significance of the social acceptance of teletransportation as survival. So, we can perhaps envision a social order in which we could find psychological strength to convince ourselves that as long as we still get to sit down to family dinners, talk about the past, raise children, and so on, the idea of a normal biological cause of the relations that constitute my psychological continuity may not be as important, in contrast to the idea of uniqueness of the survivor that was threatened in fission. These remarks are not meant to be decisive, but instead to provide an illustration of how our reflections on these issues would proceed.

The second notion of fundamental importance that seems to be threatened by teletransportation is the notion of death. The idea of mortality is among the structural bases of our understanding of personhood and its value. As such, to make the story of teletransportation intelligible as the story of persons recognizably like ourselves, we have to wonder not only about the degree to which various psychological connections hold in this case, but also about what further modifications to our self-understanding in these further respects teletransportation may introduce; i.e., we have to think precisely of the interactions between different elements that constitute our form of life. I think a very *natural* question arises here: If the scanner can record and send the information based on which your replica is constructed, then could we "store" that information to "restore" the original yet again, just in case the old original or replica dies? More generally, what is the status of such a "recording"? Reflecting on questions like this, we may end up in disagreement. Some will think that the story of teletransportation is as incoherent as

[18] Parfit acknowledges this in his section "Is the true view believable?"

fission. Others will think that there are ways to understand teletransportation as survival, as long as we can deal with our worries about the status of the recording.

Since my proposals are methodological, it is not my purpose in this chapter to resolve this disagreement or provide a definitive treatment of the issue. I think the difficulties with actually spelling out all of the subtle connections between mortality, social order, technological possibilities of replication, and so on, should be apparent based on my sketchy remarks. So, in the absence of more detailed work on these issues, we can only hope for more clarity in the future. Is the lack of resolution at this stage a problem for the proposed model of thought experiments, however? No. To tie the success of this model to its being able to give definite answers in each case misses the point of the overall aim of my project. While inconclusiveness and disagreement may doom the model that ties our judgments of survival or demise to the presence or absence of some individual features of our lives, it does not undermine mine. Discussions of the coherence of our visions of such an important issue are bound to get messy. Moreover, the fact that there may be deep disagreement in our reactions to the teletransportation case, as opposed to fission, is an important result. I have earlier indicated the clusters of features surrounding the notions of death and normal cause as the ones that have to be considered in any further story-telling about teletransportation. If that is on the right track, the overall disagreement in reactions to teletransportation may track different understandings of these notions. This thought experiment then reveals the particular conditions of our continuation for further philosophical exploration and clarification. At the end of the day, however, we may have to live with a plurality of understandings of these features of our lives.

One thing that we keep seeing, however, is that the requirement of the uniqueness of the continuer plays a very prominent role in both cases, and we will return to this point in later chapters. In fission, this requirement shows up in our inability to apply familiar concepts to the lives of post-fission products, signaling that the world of fission is different, in ways we cannot comprehend, than the world in which we can recognize persons as we know them. This is because in the fission world the entire background of the intelligibility

of our notions has been compromised. In teletransportation, however, we are moved to reflect on the significance of the normal cause of our psychological continuity, and also about death, the length of a life-span, and so on.[19]

5.5 Conclusion and Some Objections

Building on the work done in the previous chapters, I suggested that developing stories of fantastic transformations can show us salient features of our personhood. Many mainstream thought experiments already mention the details of the possible lives of persons who undergo fantastic transformations. However, the presentation of such details is abstract and shortened because such details are assumed to play a very minor role in our understanding of personhood. My approach brings to light the work that the articulation of the background of each possible world may do in outlining the constraints on the intelligibility of a vision of a possible life as *ours*. Through such articulation, we make ourselves aware of the complicated interactions between different aspects of our lives. For example, when two fission-products compete to pick up the life where the original left off, we see the limitations of thinking that we can just divide everything up (divide the life up) and solve the problem by brute force, which is what Parfit seems to imply in various remarks. Based on our reflections in the previous section, such a resolution of the problem does not hold up under scrutiny. The space of interpersonal relations, for example, is not simply conventionally adjustable to accommodate another copy of a father, brother, or friend. Our whole way of life is predicated on the presupposition of a unique continuer of a life. The difficulty of applying personhood concepts to duplicating individuals shows that it would be a mistake to classify the individuals who fission or teletransport (with the discussed caveats) as persons in our sense.

We should be careful here not to overstate the conclusion of our speculative thinking

[19] As I said, these features, manipulated by the thought experiment, may not immediately compromise the intelligibility of this scenario, which calls for further reflection. The inconclusiveness of this result may, in fact, show that there may be deep disagreements in our culture about this. Using thought experiments in the way I am suggesting may reveal this fact. Whatever is the case, getting at such problems requires envisioning the concrete details of the lives of the characters in thought experiments, which, as my approach suggests, can secure the role of thought experiments.

about the fission-survivors. Someone might reply to the whole approach that what we are testing are just the limits of our imagination. Surely, it will be said, you cannot project yourself into the world of fission, but a hundred years ago you may not have been able to project yourself into the world of *Second Life*, genetic modification, or gay marriage. Our current limitations with respect to imagining how the background world could change to accommodate the possibility of fission does not mean that the background and our practices could not so evolve and preserve 'us.'[20] The objection can then be broadly described as the threat of arbitrariness.

This is a serious challenge, and here I can only outline a possible line of response. My approach indicates that thought experiments are not infallible. This is to be expected because of the openness to further developments that I endorse. But I do not think that this feature of the approach implies that discussions of this kind have to degenerate into mere intuition-bashing. While all of the features of our life are contingent, there are degrees of importance among different features and stable patterns of interdependence among different features of the way we live that we can gauge with the aid of thought experiments. For example, there is an explicable difference between saying that what one wears, or the number of children in the family, is not relevant to the question of survival in teletransportation, and saying that the normal cause of psychological continuity is relevant. These differences can be explained by elaborating the cluster of other features of the possible world that would depend on each of these features. As I indicated in the discussion of teletransportation, threatening the requirement of normal cause has a clear connection to such fundamental notions related to personhood such as 'value,' 'death,' 'genuine memory,' and so on. The individual that emerges from teletransporter raises deep issues about the status of memory and massive delusion, and brings up questions about ordinary life-span. The conceptual reach of these issues extends to the depth of our assumptions about what a normal life looks like. Modifications with respect to any of these parameters of our lives have an important bearing on the discussion of

[20]Thanks to Marya Schechtman for pressing me on this point.

personhood. On the other hand, how one clothes oneself, or whether one has children, does not have significant conceptual implications for the issue of individual survival. In short, the patterns of dependence between different features of our life that emerge from the discussion of these differences in conceptual reach allows us to differentiate between what is more or less arbitrary, and what is contingent, but not merely arbitrary.

Now, in claiming that this difference is important and decipherable, we are of course still trading on our reactions to the cases, and we are of course dependent on shared sensitivity (sensibility) to these issues. One might say then that whatever ranking we may get of these reactions, as suggested above, they are still just our reactions, and reactions are too shaky a foundation for philosophy. In reply, I want to emphasize that these reactions are measured against other features of our lives, probed by their further developments, and explored for coherence and consistency.[21] While all of them are contingent and in some sense must depend on various sensibilities that members of our community share, they are judged by accepted standards of reasonable reflection. These standards are surely fallible, but they are fallible when we reflect on anything at all, and not just in these cases.

Still, what is this contrast between intelligibility and unintelligibility to which I am appealing? This requires more spelling out, which is outside of the scope of this project,[22] but here is a very brief attempt. In order to understand whether the survivor of some fantastic ordeal is the same person who was there at the beginning, some background of intelligibility has to be presupposed. For example, to understand purposeful action, one has to assume some basic facts about rationality, embodiment, the structure of intentional action, and so on. Similarly, if we are interested in the question of survival (and other questions of personal identity), we have to presuppose some 'metric of persons.' In fission, this metric is under threat, thus compromising our judgments that only make sense within it. But then it may be that the traditional questions of personal identity do not apply here. These questions are presumably generated by an interest in accounting for phenomena of this world, and the expectation that we can rely on familiar conceptual features to address

[21] Think of Rawls's (1971) reflective equilibrium here.
[22] Brandom 2008 may be seen as one such attempt.

these questions. As I indicated earlier, one of the starting points for many philosophers interested in the metaphysics of persons was to probe the justification for our practical concerns. Now, as we are trying to critically engage our construction of the world of fission, much of what gives sense to the question of personal identity as relevant to ourselves in the conceptual space of familiar practical contexts is simply not in the picture. The structure of the consciousness of fission-world inhabitants will surely be different from ours. Our questions may not apply to it.

I am aware that this response does not address the worry conclusively: there is an unavoidably subjective element to our imaginations and their failures. But there may not be any other way to go on with the cases, apart from our continuing case-by-case reasoning about fantastic situations, to get to some possible agreement about them. These agreements can only come from further reflection not merely about our reactions, but about the way that they fit into our broader range of interests in the guiding questions of personal identity as relevant to practical concerns.

Chapter 6

Divide and Conquer

6.1 Introduction

In the previous chapter, I applied the new model of thought experiments to two famous cases: division and teletransportation. I argued that by expanding the range of considerations relevant to our assessment of these cases to include the context of our practices and the institutional background of our lives, we can gain insights into complicated relations between different elements of our lives that constitute the conditions of intelligibility and the material conditions for our continuation. In general, our discussion so far proceeded under the assumption that practical concerns have a direct connection to metaphysical questions of personal identity. Indeed, one of the traditional motivations for exploring metaphysics of persons was to provide justification for our practical concerns (Shoemaker 2007). However, as I indicated in various places, this presumed connection is not universally accepted, and both metaphysicians and ethicists have recently been questioning it.[1] In this chapter, I address the more radical version of this challenge. If this objection succeeds, it compromises the usefulness of the literary model of thought experiments I am defending. The challenge is to show that the details of practical lives are directly relevant to the metaphysical question of personal identity.

[1] In general, this has been one of the more exciting recent developments in the literature on personal identity.

6.2 Practical Concerns and Personal Identity

At this point, the literature exploring the critical connection between practical concerns and metaphysics of identity is voluminous. A review of this literature is not essential for my project, as long as we can briefly articulate the presumed connection. As David Shoemaker (2007) observes, many philosophers come to be interested in the metaphysics of identity, and specifically the question of diachronic identity[2] because it presumably provides rational grounding for our practical concerns (317). It is natural to presume that answering such questions has to do with identity. Thus, we presumably hold somebody responsible for some past action because she is the same person who acted in the past; we compensate somebody for past burdens because she is the person who sacrificed so much for the achievement, and so on. Consequently, it is natural to assume that the answer to the philosophical question of reidentification—is the individual p1 at t1 the same person as p2 at t2?—will help us settle questions arising about our practical concerns. Moreover, because practical concerns, on the face of it, require psychological continuity between persons across time, criteria of identity that do not make reference to psychological relations have been assumed to be implausible. This assumption has tended to favor the psychological continuity theory of identity over its rivals.

Thought experiments have played a significant role in various defenses of the psychological continuity theory. In most ordinary cases, it may seem that biological continuity can supply as good an answer to the philosophical questions of identity by assuming the link between psychological continuity and its physical realizer, the brain. The assumption is that in thought experiments we can separate the two to get the proper judgment of what is required for the preservation of identity.[3]

[2]The question is what makes a person p1 at a time t1 the same as the person p2 at a time t2?

[3]As an illustration, consider Locke's classic case of the Prince and the Cobbler, in which the soul of the Prince "enters and informs" the body of the Cobbler. Locke says that anybody would judge that the person Prince is whoever assumes the responsibilities of the Prince, because that person has memories that belong to the Prince. In this case, we are clearly assuming—at least Locke thinks that we are—the link between practical concerns, such as responsibility or compensation, and identity. What happens to the old body of the Prince is not relevant to the question about the person Prince. As we saw in the previous chapter, Parfit's own relation to the question of the connection is more complicated: after all,

The central assumption that practical concerns and the metaphysics of identity are inherently connected has recently come under increasing pressure from different sides of the debate. The details of these criticisms are not as crucial as the general shape of the objection. The objection—which I will call the 'divide and conquer' strategy—urges that practical concerns, including ethics, cannot play the decisive role that they have been playing so far because practical concerns do not have to track personal identity. In what follows, I focus on Eric Olson's defense of animalism and Susan Wolf's argument against Parfit's conclusion that identity is not what matters in survival. After that, I sketch a response to the divide and conquer approach, based on Marya Schechtman's work. My purpose here is only to indicate that there is logical space for some metaphysical question of identity that is intimately tied to our practical concerns, and thus to make room for an understanding of thought experiments according to my model. Providing a positive account of such a metaphysical entity is beyond the scope of this project.

6.3 Divide and Conquer: Olson

The appeal of the psychologically-based accounts of personal identity lies in their promise to ground our fundamental practical concerns. However, Eric Olson (1997) argues for the conclusion that psychological relations—the relations that are typically directly connected to our practical concerns—have nothing to do with the metaphysics of identity. According to him, 'being the same person' is simply not a metaphysical relation, even if we find questions about this relation to be of overwhelming importance (Olson 1997, 69).[4] This view is diametrically opposed to the central assumption of much of the literature on personal identity. Let's look at some of the details of Olson's position that are necessary

he argues that identity is not what matters in survival and in other practical concerns. However, what matters is relation R, defined in terms of the holding of the relations of psychological continuity and connectedness. Having given up on identity, then, Parfit has not given up on the idea that practical concerns can be used in investigations about who we are. Sameness of person is still defined in terms of relation R, and Parfit uses thought experiments that invoke judgments about practical concerns to investigate the behavior of this relation.

[4]He writes: "Ultimately it is for ethicists to tell us when prudential concern is rational, when someone can be held accountable for which past actions, and who deserves to be treated as whom. These are not metaphysical questions because, *being the same person*, as we might say, is not a metaphysical relation."

for understanding the general shape of the objection I am exploring here.

1. Olson points out that there is an ambiguity in the way the question of persistence has been raised. Typically, the question of personal identity has been formulated like this: what is it that makes person p1 at time t1 the same person as person p2 at time t2? According to Olson, this formulation privileges psychologically-based accounts of identity by seeking the connection between two persons, or person-stages, in terms of the relations of psychological continuity. According to Olson, we can ask a different question: what is it that makes an individual that is a person p1 at time t1 the same thing that is a person p2 at time t2? This formulation, according to Olson, is neutral with respect to what kind of account of persistence we give because it does not confine us to a psychologically-based account.

2. At the heart of Olson's challenge to the defenders of psychologically-based accounts is the idea that, in order to address the metaphysical questions of identity, we need to ask in the most basic sense, *what* kind of thing we are dealing with; i.e, we need to ask about its 'substance concept.' A substance concept "tells us, in a special sense, what the object is and not merely what it does or where it is located or some other accidental feature of it," and it is the substance concept that determines our persistence conditions (Olson 1997, 28; DeGrazia 2005, 28). An entity can fall under only one substance concept, and an entity cannot fail to fall under its substance concept without ceasing to exist. Contrasted with the substance concept is a "phase sortal"—a temporary kind to which an existing thing can belong or not belong at different stages of its substance-kind career (Olson 1997, 28; DeGrazia 2005, 28). Temporary kinds are, for example, barrister, mother, professor, student, and so on. To illustrate the difference, consider that when an individual becomes a student, she does not go out of existence, while the new thing—student—comes on stage. Instead, the individual in question enters a new stage of being a student, which she will leave, hopefully, at some later point without thereby ceasing to exist. At the same time, according to Olson, individuals like us cannot stop being biological organisms without going out of existence.

3. Olson claims that our substance concept is 'human animal;' defenders of psychological continuity claim that our substance concept is 'person.' What drives the idea that 'person' is our substance concept, or the idea that our identity is tied to psychology? Well, it is various intuitive reactions that we have in response to puzzle cases. (Consider Parfit's discussion of teletransportation and fission, covered in Chapter Five of this thesis.) On the neutral formulation of our starting question, however, the space is open for any account of continuity. Thus, our intuitive reactions tied to psychological continuity are not necessarily tied to answering the question of identity.

4. In fact, Olson uses Parfit's argument that identity is not what matters in survival to drive a wedge between the psychological continuation (related to practical concerns) and the relation of numerical identity (metaphysics). Since they can go their separate ways, Olson argues, our intuitions in puzzle cases are compatible with animalism as a theory of numerical identity, and after Parfit this should not be surprising.

5. There are good arguments to think that 'person' makes for a bad substance concept. First, psychologically-based accounts of personal identity have difficulty accommodating what seem like accepted facts about ourselves. For example, we accept that I was once a fetus. However, a psychologically-based account defines my persistence conditions in terms that are simply not applicable to the fetus. Then either I was never a fetus, or there is something missing from the psychologically-based account. Another problem is that the psychologically-based accounts have a hard time explaining the relation between the person and the animal. For example, there seem to be two individuals writing this line: the person and the animal, and it is not clear which *I* am.[5] Moreover, 'being a person' has features of a functional kind rather than of a natural kind. According to Olson, being a person is just a set of capacities that an animal may possess: thinking, self-consciousness, and so on.[6] So, if it is shorthand for a bunch of capacities, Olson holds, being a person is just a stage in the career of an organism that occurs when these capacities develop.

[5]See Baker 2002 for a response to these objections.

[6]DeGrazia's list includes the following capacities: autonomy, rationality, self-awareness, linguistic competence, sociability, the capacity for intentional action, and moral agency. DeGrazia (2005) says that a person is someone with the capacity for complex forms of consciousness (6-7).

Like an office, one can come to occupy it and may leave it earlier than at the point of biological death, as happens, for example, with people who enter a persistent vegetative state (PVS) (Olson 1997). Lastly, saying that we are persons rather than animals posits a break in the biological continuity between human and non-human animals.

To illustrate these points, consider an individual in PVS. According to Olson, a metaphysician's question about this case is whether you are still there, "in any state at all," and not whether you are thinking, feeling, or are conscious. He writes:

> If you had said, before the accident, "I shall one day be a human vegetable, lying unconscious on a hospital bed," would you have said something true? Or would destroying your higher cognitive functions bring your existence to an end, just as any ordinary means of death would?.." Do you come to be a human vegetable, or do you cease to exist and get replaced by a vegetable, much as you might be replaced by a statue? (Olson 1997, 9)

I admit that there is something right about the idea that there is *some* very real sense in which the individual survives the loss of all her cognitive functioning. Here is the living thing you were throughout your career in PVS, and it is not dead; it is just that this individual is not thinking or talking because he has lost some of its many capacities. Based on such considerations, Olson draws the conclusion that it cannot be the psychological relations that define the continuation of this individual: they are simply not there, even though the individual has not gone out of existence. So, such relations are in fact "completely irrelevant" to identity (Olson 1997, 20).[7]

Given that Olson appeals to rather ordinary considerations, how do we explain the appeal of the psychologically-based theories? According to Olson, this appeal comes from the overwhelming significance that practical concerns and thus the practical sense of 'being the same person' play in our lives (1997, 65–70). Ordinarily, the practical concerns coincide with personal identity, and it would be easy to confuse tracking intuitions about one (practical concerns) as determining the answer to the other (metaphysics). So, confronted with fantastic cases, we mistakenly think that the relations that underlie responsibility,

[7]On the other hand, if you thought that 'person' was our substance concept, the individual in PVS has come into existence at the moment when the person's psychological features collapsed.

compensation, and social treatment should inform our answers to the metaphysical question of identity. As Olson shows, this is a confusion. As long as the animalist can explain standard puzzle cases as being about this practical sense of being the same person, he can claim that metaphysical investigations do not have to be tied to practical concerns. If such an argument is convincing, our intuitions in thought experiments do not track what we most fundamentally are.

6.4 Divide and Conquer: Wolf

Susan Wolf can be interpreted as advocating an approach that is structurally similar to Olson's, despite the differences in philosophical sensibilities and the details of their argumentation. Let's discuss her view in some detail.[8]

Recall what I said earlier about the prospects of fission-products. Their lives, according to Wolf, will be much worse than in ordinary survival (1986). Wolf may have shown that the consequences of fission for some lives may make ordinary survival better than surviving division. But, as Parfit claims responding directly to Wolf's paper, in *his* case he stipulated that the prospects of the survivors are as good as they would be without division. He writes: "To block this argument Wolf would need to show that there cannot be such a case—that division would itself ensure that the two resulting people would have prospects which are worse than mine. This I believe she has not shown, and could not show" (Parfit 1986, 864). The shape of Parfit's argumentation is as follows. Wolf must argue that the quality of our prospects in survival is tied to the idea that there is a "deep further fact" of identity. Thus, if there is no identity, the prospects must be bad. If this is a general claim, all we need is one counterexample to this idea. But we can just stipulate such an individual with fission-prospects that are as good as ordinary survival: his personal relations are dispensable or replaceable, his income can be

[8]Here, I use Wolf as a representative of such an approach, but of course there are others. Christine Korsgaard (1989, 2009), whose views I discuss in the next chapter, may be seen as advocating a similar division of labor between metaphysical considerations and practical concerns. According to her, there is no reason to think that we are going to find grounding for our practical concerns whose natural home is in the practical standpoint in the considerations about metaphysical composition of entities, whose home is the theoretical standpoint.

divided in half without much problem, he does not care where he lives, he does not mind abandoning his long-term projects, nor does he mind that there is another person who can legitimately claim to be the authentic survivor, and so on. In short, "[his] relation to each resulting person contains everything that matters in ordinary survival" (Parfit 1986, 863). This counterexample shows that there is no conceptual tie between division and one's prospects such that the prospects of division are worse than in ordinary survival. This explains why Parfit freely admits that in some cases division is worse than ordinary survival. Furthermore, by Wolf's own admission, her imagination may have not been sufficiently open to appreciate that the positive change may be our becoming less focused on ourselves while becoming more concerned with others (Parfit 1984). So, it may seem that the arguments are on a par.

But there is a more general line in Wolf's argumentation. As she puts it, the *grounds* for answering the question of what ought to matter to us have to do with the moral and evaluative practices that constitute our form of life (Wolf 1986, 713). Whether identity matters depends on how it hooks up with the rest of the evaluative practices we have (Wolf 1986, 714).Wolf herself admits to being convinced of the metaphysical truth of reductionism, but according to her, Parfit goes wrong in assuming a rather simplistic relation between the metaphysical truth and the question of what ought to matter to us. What ought to matter to us must start with reflection on what is valuable about persons as we know them, and this involves us in considerations that have little to do with the metaphysical composition of persons (Wolf 1986, 708, 716). Instead, we should compare whether caring about persons as we do now is better than caring about Parfitian persons. She says:

> [w]hether personal identity matters depends on how it connects to other things that matter. But, as we have seen, the other things that matter, the things having to do with the quality of life, or the quality of the form of life, are themselves features of what we may call the surface of the world. Their value is independent of their metaphysical composition."(Wolf 1986, 713–714)

While it may be that there is no "further" metaphysical fact of identity, the value of

uniqueness is located in what Wolf calls the surface level of value.[9]

This point is important. Parfit assumes that the truth of reductionism, removing the veil of misguided attachment to identity-based considerations, should result in ethical revisionism. Seeing that identity is just a subspecies of R, Parfit suggests that our concerns, presumably based on identity, should shift to reflect the features of R: it can hold to a degree, and it is not unique. But Wolf argues that our concerns and the things we value do not simply come and go with metaphysical (or scientific) understanding. Moreover, saying that our institutions will adjust to reflect the newly discovered truth of reductionism does not undermine Wolf's point. Suppose we think, on the basis of imagination, that the possible worlds in which we split are better than this one. We still have to explain why we should take *that* fact to be of significance to us, persons who do not split. And to answer this question about the reason for caring about the comparison we again have to appeal to our values and attachments. Thus, even though we will be considering worlds with different metaphysical compositions, so to speak, the evaluative import of the connection has to do with non-metaphysical issues.

Parfit (1986) responds that Wolf's argument is *incommensurable* with his (864). One can say that Wolf's appeal to consequences comes from within the current framework of value and significance, based on the assumption that there is a deep difference between being me and being someone else. But since Parfit is purporting to *see through* the veil of identity, his argument does not have to share in Wolf's presuppositions of how we currently approach value—after all, according to him, since there is no deep difference between being me and being somebody who is very much like me, much of what we value is the product of an illusion.

However, I take it that Wolf's point is that Parfit himself must rely on the entire web of values and concerns of a given individual (like himself), for whom duplication can look as good as ordinary survival (in his case). So it seems that, contrary to what Parfit

[9]Wolf concludes: "Whatever significance we assign to the question of whether the altered world would be better than ours is, in any case, independent of the metaphysical issue of personal identity. The truth of reductionism neither detracts from nor supports the reasons advanced for thinking that personal identity matters" (1986, 716).

thinks, what is doing the work in his argument that duplication is as good as ordinary survival is the individual's life-circumstances and values, and not only the recognition of the metaphysical truth about identity. To say whether duplication is worse, better, or equal to ordinary survival, one has to do more than count the strands of psychological continuity; one has to say—as Parfit does—that nothing is missing in such survival. Thus, one has to take into consideration the place and significance that the number of strands of psychological continuity and connectedness between the original and the fission-products plays in the overall web of our concerns. This is enough to make Wolf's point, I think.

Suppose so. Still, doesn't Parfit's case actually refute the idea that metaphysical identity is what matters? Look, one might insist, the individual survives without identity, doesn't he? However, Wolf's point is that *if* you think that the relation between mattering and identity is that the metaphysical matters settle the question of what ought to matter, then you *may* be right. However, the relation between mattering and identity is not something that you get out of the argument itself. It is a presupposition that may not be shared by all parties.

While there is a great distance between Olson and Wolf regarding many philosophical points, they share the idea that the metaphysics of identity can be pursued independently of our practical concerns. This compromises the appeal of my model of thought experiments. As we will see in the next section, however, there is a response to this strategy.

6.5 Response to Divide and Conquer

The divide and conquer strategy pursues questions about metaphysical units found in the world (individuals that fall under substance concepts, reductionist bits of strands of relations of psychological connectedness and continuity, and so on) separately from questions of the relations between the metaphysical units, including the questions of value that arise from such relations. Perhaps we can make some sense of the idea of metaphysics as completely detached from practical concerns: as an investigation of abstract

principles of composition, for example.[10] However, it goes against the intuitive and natural appeal of the idea that facts about personal identity are intimately connected to our practical concerns—the starting intuition behind traditional approaches to personal identity. Against the sharp separation, I will argue that the questions traditionally asked by Locke, Parfit, Lewis, Perry, and many others, are not on the same level with questions about the details of ascriptions of responsibility, compensation, and so on, in some particular set of circumstances. Following Schechtman (2008), we can distinguish between direct questions of value and the conditions of asking such questions. I will call the former 'internal' and the latter 'external' questions about our practices.[11] The internal questions are about the circumstances in which this or that person is rightly blamed or praised. The external questions are about general conditions of there being a practice of praising and blaming at all. According to Schechtman, at least some of the traditional puzzle cases of philosophers are about the fundamental *preconditions* of the discussions of the details of the practical concerns themselves. As such, they are meant as questions about the proper basic unit of practical concerns.[12] Unfortunately, the two sets of issues get confused, as witnessed by the fact that the "divide and conquer" strategy puts all questions that have to do with our more general evaluative practices on the same plane. Maintaining the distinction paves the way for a response to the challenge. If this response is successful, there is room for an alternative understanding of a metaphysical question of identity that is directly tied to practical matters, whatever strictures one might stipulate to ensure that that understanding is properly austerely metaphysical in Olson's sense.

Fulfilling this task will first require looking at some of the details of Olson's argumentation, which targets the psychological continuity theory. This discussion will put us in a better position to appreciate the response to the "divide and conquer" strategy.

[10]The connection is drawn in Olson's *What Are We?* (2007). I am not confident this separation is as clean as one may want. For instance, I agree with Korsgaard, who rightly points out in footnote 47 of her reply to Parfit (1989) that countability and ontological economy, the traditional concerns of metaphysics, are just some concerns among many. So, calling those 'neutral' is question-begging.

[11]Associations with the views of Rudolph Carnap are not to be taken too seriously here.

[12]It may not be exactly what Olson meant by his question of identity "in any sense at all", but neither is it just about the particularities of the practices that are the stuff of life of such units.

One of the standard arguments in favor of the psychological approach in personal identity is the prevalence of intuitive reactions that biological continuity is not necessary for the survival of the person. As you recall, Olson's strategy is to show that there is no incompatibility between animalism as a metaphysical theory of identity and our intuitive reactions triggered by the favorite cases of psychological continuity theorists. Olson (1997) discusses three different practical concerns that, according to him, we mistakenly assume to be connected to numerical identity: prudential concern, moral responsibility, and coherent social practices (52–70). As Olson points out, many mainstream psychological continuity theorists are themselves perfectly happy to sever prudential concern and identity. Parfit's famous fission case discussed in Chapter Five amounts to just that: (numerical) identity is not what matters in survival. It is then perfectly legitimate for Olson to ignore the intuitions about such concern as irrelevant to the metaphysical truth about identity (Olson 1997, 56).[13] Thus, animalists can say that when the "soul" of the Prince enters and informs the body of the Cobbler, the fact that we would all say that the Prince goes where his psychology goes is no indication of some metaphysical truth; it is, rather, a case of the profound transformation undergone by the Cobbler's organism by way of getting a new cerebrum. Moreover, by looking forward to his continuation in the new body, the original Prince is not tracking any metaphysical truths about himself. His concern, according to Olson's elaboration of Parfit's view, does not have to be a concern for himself, but for someone else appropriately connected to him via the relations of psychological continuity—via being in a new 'office.'

As Schechtman (2008) notes, however, these cases should not look so similar on closer inspection. Consider the following two cases: fission and the Young Russian. We have seen that the fission case results in the double transfer of the psychology into the bodies of your twins by means of some amazing memory-transfer device. In contrast, in the case of the Young Russian Parfit asks us to imagine something more mundane—that the noble Russian youth, anticipating his future changes, asks his wife to promise him that

[13] Also see a similar move in DeGrazia 2005, 62-64.

whatever he says in the future, she should honor the promise to his prior self. In this case, we are talking about the diminished psychological connectedness between the current and the future person-stages of the Russian (Parfit 1984, 327).

Speaking very generally, both are cases of *change*, and at this general level, a relocation of a whole mind is no different than a change in the contents of a given mind. As Schechtman (2008) argues, however, there is a conflation here of *transfer* cases with cases of *continuation*, and this is an important difference. The spectrum cases, according to her, exhibit this conflation particularly well. Recall that Parfit's (1984) spectrum cases picture a series of experiments in which a surgeon at each step flips a switch that preserves fewer and fewer psychological, biological, or both types of connections with the original (231). For example, one can imagine a spectrum of cases of varying degree of psychological connection between Parfit and Greta Garbo: at the beginning, there are few psychological continuities between the two, while by the end, not much is left from Parfit's psychology, and the new person is mostly like Garbo. According to Schechtman, in the psychological spectrum, for example, while all the way up to the very last case we have continuation cases, with the last step the surgeon causes all the connections between Parfit now and some future person to cease, effectively transferring a new psychology (Garbo's) into his body. Since the last case in the spectrum is presented as just another step in the sequence, it looks like what is essentially a transfer case is the same kind of case as the regular psychological changes at the beginning of the spectrum, which in turn is the same kind of case as the continuation cases of the Young Russian and the rest (Schechtman 2008, 43).

One might say that of course the cases are of the same kind because all of them are about relation R: from strong relation R to none. According to Schechtman, however, this conflation of cases obscures an important difference. Let's discuss the difference in detail. When we think of regular cases of gradual psychological changes, like that of the Young Russian,[14] we can certainly get on board with the idea that being a particular person with

[14]Olson discusses an example of Milan Kundera's character Tereza, who wonders whether she is the

particular character traits and memories is not something that is guaranteed to be stable throughout the career of an individual. Cases of this kind, which involve profound changes in the content of psychological relations, are useful in exploring questions of responsibility or compensation. For example, when an older person looks back at her past, reflecting on all the changes that have occurred, she may not at all feel anything for the person who she once was. Or the Old Russian can look at his former self's wishes as naive and pointless. And third parties will take these alterations in assigning blame and praise. Such cases are useful in discussions of identification in the sense of reflective endorsement.[15]

But the transfer cases surely *seem* different from the cases of Tereza and the Young Russian. The relation 'being the same person' may be taken in different senses. Tereza may say that she is now a different person than she used to be because she does not believe the same things she used to, does not care about the same issues, cannot see her current self in such and such ways, etc. In contrast, when we turn to the standard puzzle cases like teletransportation and fission, it seems clear that the authors of these thought experiments are not interested in assessing the stance with which the current person looks at her past or future. It is stipulated that there is full psychological continuity between the past and the future person; there simply is no dramatic personality change involved in teletransportation or in fission. Instead, these cases are used to explore the issues connected with the cause of the continuity, or the uniqueness requirement, and so on. Locke, Parfit, Perry and so on, are interested in a different sense of 'being the same person'—the more general question of what it is that we are prepared to call the kinds of continuation that preserve persons. Suppose Tereza is about to step in the teletransporter. Her worries, it seems plausible to say, are not specific worries about whether her continuer will identify with what she identifies with at this moment. Instead, Tereza's question is this: will the procedure destroy *her* or not? It is about the general conditions of continuation of persons; if you will, it is about the causal structure of the

same person as her younger self given how little of what she used to believe as a young woman she now thinks is true (Olson 1997).

[15]These senses of identification are explored by Korsgaard 1989, among others.

world.

So, while it is stipulated that the concerns and worries will be preserved, from both the standpoint of the individual who appears on Mars and the standpoint of somebody who is about to be "beamed" it remains to be seen *whose* worries they are—whether they are *attributable at all* to a given person, as opposed to the more specific question of *the degree of attribution*. To determine whether my replica on Mars is responsible for my actions here and now, we first need to show that whatever it is that is necessary to begin to attribute responsibility to an individual at all—call it sameness of individual consciousness, in Locke's sense, for example,—is preserved through the operation. If it is not, then it is problematic to claim that my replica is responsible for my actions—not because my replica fails to identify with the person who performed these actions (because we assume that the replica has the same psychological dispositions), but because my replica does not stand in the right kind of relation to the action which is being attributed to him. (The treatment will be different in the fission case, of course.)

Of course, we can insist, with Olson, that any question that involves us in the discussion of practical concerns puts us outside the metaphysical discussion, but I think such *foundational* or *criterial* questions can be distinguished among the more specific ones. We can put it this way: when Olson claims that it is ultimately for ethicists to tell us about what is required for an assessment of the rationality of prudential concern, or in deciding who is responsible for whom, etc., the phrasing is equivocal. On the one hand, we may only want to show that responsibility requires the relation of psychological continuity, or that prudential concern may be grounded by a relation other than the relation of psychological continuity. But addressing such questions requires that we presuppose some stable facts about the world such as that people do not switch bodies, do not regularly abruptly change their deeply held moral convictions, cannot upload or unload their psychologies, to mention just a few. Such presuppositions form the background of our lives and anchor the particular discussions. In the standard thought experiments in personal identity we are exploring (questioning, probing) changes in such basic facts rather than changes to a

given person's psychological features.

We can quibble about whether these questions are of interest to metaphysicians, but it is uncontroversial to assert that exploring the preconditions of our evaluative and normative practices is *at least* at the borderline of ethics and metaphysics. The distinction between the questions about the particulars of our practices and the proper units about which such questions can arise paves the way for a response to Olson.[16]

Now, whether or not the proper unit of our practical concerns is also a proper candidate for being a substance term in Olson's sense is a further question. So is the question of the relation between the animal that we are and the proper unit about which we can ask our questions of practical concern. While I won't be addressing these further questions, we should be aware that there are various options on the table. First, there are powerful responses to animalism from the constitution view of personal identity (e.g., Baker 2002). The view claims that we are fundamentally persons constituted by animals. So animalism is not the only theoretical option, and it may be that there is some affinity between the view I am putting forth and the constitution view of persons. Additionally, as we will see in a moment, Schechtman has been arguing for an account different from animalism of the proper metaphysical unit about which the questions of practical concerns can arise, namely the person-life view. If she is correct, then the question of the relation between the proper unit of our practical concerns and our animal will have a different status than it does on Olson's theory. For my current purposes, it is enough that the presence of the on-going discussion about this issue makes the field open for the kind of engagement with thought experiments that I advocate. Finally, if one were to deny that this is metaphysics, properly understood, then I am not doing metaphysics. However, I am not alone in this departure.

This response cuts against Wolf's version of the division of labor approach to the issues as well, but slightly differently. There are two things that Wolf says about the relation between the metaphysics of identity and the question of what matters, strictly

[16] Again, I am deeply indebted here to Marya Schechtman's work.

speaking. On the one hand, she claims that reductionist considerations neither support nor detract from our conclusions about what matters; on the other hand, she mentions that whether identity matters or not depends on how it *hooks up* with the rest of the evaluative practices from the surface of this world, so to speak. If we focus on the second point, considering it in the context of our discussion, we may come to a more nuanced understanding of Wolf's position.

One way to picture the dependence between the two is not to insist that the metaphysics of personal identity determines the behavior of our practical concerns because it justifies them (so that absent the justification, the concern is compromised), but rather to think of metaphysical facts (not restricted to facts about compositional simples) and practical concerns as intertwined and mutually dependent.[17] Of course, Wolf would not deny that there are certain natural facts about us that make, for example, our evaluative concerns and our practices in general possible. Such facts do not have to directly justify each and every practical concern we have, even though their presence shapes in different ways our practices and the possibilities of their future development. Thus, while Wolf denies that the truths of metaphysical composition are relevant to our value considerations because the latter are independently articulable and justifiable by what is properly called the surface of the world, I don't exactly deny this, but I think we have seeds for thinking about other kinds of dependences between values and things in the world. It would be strange to think that there is not some connection there. The proposal is then to broaden the options of what the relation between the two may be.[18] In addition, as I have been arguing in the previous chapters, the practices we have are not static, and we should expect that future technological developments, for example, will present us with opportunities to reassess what we value by virtue of remapping the space of possible physical modification a body and a life can undergo. (See Chapter Three.) Some of the things we value are easier to modify than others, and this has to do with the complicated

[17] In Chapter Three, I have given some examples of such co-dependence.

[18] To see this point, consider the stalemate between defenders of sparse ontology and Baker's (2008) 'big-tent' metaphysics, which sees ontological questions as dependent on what we value.

relations between facts, our concepts, and questions of value.

6.6 Response to Divide and Conquer and the Literary Model of Thought Experiments

In the previous section, I provided a reply to Olson's charge that defenders of the psychological continuity theory are guilty of confusing ethics and metaphysics. In response, I followed Schechtman's distinction between direct questions of value and questions of condition of value to generate the view that we can distinguish between two different ways in which practical concerns can be used in the discussion of persons. On the one hand, we can be interested in the question of the degree of attribution of a particular concern within the practice; on the other hand, we can use our intuitive reactions about practical concerns to address the question of what counts as the proper unit of such concerns. This question is about the structure of the framework of our person practices, rather than about the details intrinsic to the practices. In the context of this dissertation, drawing this distinction opens up a way of understanding how practical concerns can be directly relevant to the metaphysics of identity. If such a response is successful, I can meet the charge that looking for insights into metaphysics in the uses of imagination in literature and in bioethics amounts to changing the question.

Suppose this clears one of the general theoretical obstacles to connecting metaphysics and practical concerns. But is there a natural link between the methodological proposal of the earlier chapters—the literary model—and the theoretical discussion in this chapter? Indeed, I think the link is that elaboration of further details in thought experiments puts the psychological elements of our life back into the context of the world of the interaction of things, thus returning the question to the level of literal identity. Let me elaborate in a somewhat roundabout way.

Recall that when Locke considered the case in which the soul of the prince "entered and informed" the body of the cobbler, the difficulty for Locke was to explain what that sameness of consciousness that can be moved around would be, especially in light

of his explicit argumentation against substance–based views of personal identity (Locke 1975).[19] As Schechtman (2008) argues, Locke's failure to explain what sameness of consciousness amounts to, and the subsequent changes to his original insight that took place in Neo-Lockean developments, have ultimately resulted in various problems with the psychological continuity theory. According to her, and this is the interpretation I have been endorsing, Locke was interested in giving an account of the kind of entity, or thing, about which the questions of our practical concerns can naturally be asked. His alternative to spiritual and material substance views was sameness of consciousness: what it is to be me is to continue as the same experiencing subject. But since this is not a traditional substance view,[20] making sense of what makes for the sameness is difficult. According to Schechtman (2008), the subsequent developments of Neo-Lockean accounts added other psychological connections, in addition to autobiographical memory, introduced a causal component, and instead of direct psychological connections introduced overlapping chains of such connections (40–41; also see Rovane 1990). Locke's focus on phenoomenological unity explaines his origianl insight that the metaphysical unit that can account for our practical concerns has to be captured in terms of sameness of the experiencing subject.[21] Neo-Lockean account, in contrast, are characterized by an overly exclusive concentration on the *content* of the psychological relations that ultimately erases the difference between somebody being *me* and somebody being just *like me*. This is because relations of psychological continuity—considered exclusively on their own—can be replicated. While Locke's view began as the candidate promising to account for our practical concerns, these changes make Neo-Lockean versions of the view subject to the extreme claim.[22]

I would like to think that these Neo-Lockean developments of Locke's view are both partly reflected in, and responsible for, the abstract form that standard thought experiments take. By focusing too narrowly on the intrinsic nature of a privileged set of relations

[19] I am sure this is controversial, and Locke scholars will disagree whether this is Locke's view.

[20] However, see Jessica Gordon-Roth's recent articulation of Locke's view as a kind of substance view, just not the material or spiritual substance.

[21] Schechtman develops a phenomenological reading of Locke's view already in 1996.

[22] For full details, see Schechtman 1996, 2008.

in abstraction, and by putting to the side, ignoring, or insufficiently caring about the practical situatedness of such relations in the world, the thought experiments tend to picture consciousness as some isolated phenomenon whose interactions with the world are a kind of an afterthought. (For an example of this kind of treatment, see Parfit's evaluation of the practical consequences of fission in the previous Chapter.) But if the foregoing argumentation is correct and the original inspiration of the Lockean picture had to do with thinking about providing a criterion of identity for an entity that is defined in terms of "forensick" concerns, it is the kind of entity that is most naturally understood in the world of social interactions that are mutually intertwined with our practical concerns: responsibility, compensation, prudential concern, and so on. This social embeddedness imposes constraints that may be invisible if we focus too narrowly on just the intrinsic psychological features of the isolated individual. Thus, to recover the lost sense of the connection between consciousness and the world of these practices, I will urge the picture of thought-experimental background development as the missing element in our theorizing about personal identity.

This is where further elaborations of thought experiments come to play their role. Telling stories puts the protagonists of thought experiments in the field of action, and allows us to trace the consequences of the fantastic transformations in that field. Our world is the world of interaction with others, where you have to think of "interface" issues: issues of mutual constraining and development by the "inner" and the "outer." Because we live among others who have bodies, plans, relationships, and so on, all of these facts constrain and determine the shape of our actions, evaluative practices, psychological relations, and so on. Our actions have to be intelligible to us against the background of the overall understanding of rationality that is accepted in our culture. Richard Wollheim (1988) expressed this thought in the following way: thought experiments cannot really succeed in telling us about persons unless they put the fantastic vicissitudes inside a conception of a life that this entity leads.[23]

[23]See the discussion of Wollheim in Chapter Three. In connection to what I just said about putting

The point of mentioning these issues in our context is this. What is at issue in metaphysical discussions of personal identity is understanding persons as entities whose lives occur in the world. Both the nature of these lives and our reflection about them are conditioned by a set of biological, cultural, environmental circumstances. If, as I argued, the defenders of the psychological continuity theory are asking the question on the level of entities, then they need to consider the impact of the fantastic vicissitudes on the entire range of issues that arise for our lives considered in their natural circumstances, since these aspects are involved in our understanding persons. So, for example, if we think that part of our conception of persons is tied to the idea of living a life with a certain temporal span, then it is important to look beyond (project beyond) the narrowly focused situation that is described in the thought experiment. We may wonder, as I showed with regard to fission in Chapter Five, what consequences follow, and how, if at all, one might deal with those consequences. These kinds of investigation explore the sustainability of our notion of personhood in the possible world of fission. Or, if we think that 'person' is a forensic term,[24] following Locke, we may want to look at what fission does to challenge various practices, centered around persons, such as ascribing responsibility, making compensation, and so on. As we have seen, it may be that the consequences of fission are so dramatic that we do not know how to extend our practices to fission-products, which may in turn compromise our view that we could apply what we think about these entities to ourselves.

To avoid misunderstanding, the argument here is not simply that our concepts like 'responsibility' or 'compensation' don't apply in the new circumstances. Depending on

the intrinsic psychological relations back into their natural context of lived life, discussing the lessons we can extract from Ovid's *Metamorphoses*, Wollheim argues that vast areas of our understanding are missing when we try to understand what the guise of a non-human animal does to a person. According to Wollheim, if we do not connect the animal appearance with the features of inner life that a typical human leads, then all we have is a picture rather than a genuine possibility. To simply superimpose the human inner on the non-human outer is to ignore serious constraints that are imposed on us by their interaction—the "necessities" of life. I take it that this is also partly the lesson of Kafka's *Metamorphosis*.

Similar insistence on the importance of the background details for intelligibility, albeit in different contexts, can be attributed to at least some of those who defend narrative approaches to personal identity. For example, as McIntyre (1981) argued, for any description of an action to count as such, it has to be a description given against some background, which makes it a candidate for being understood, it has to be embedded into some kind of story, however bare.

[24]See West 2008 for a defense of a strong view of this kind.

the case, our conceptual apparatus may be flexible enough to incorporate and sustain the possible changes in the range of our powers to prolong and modify human lives. Instead, the argument is that our thinking about the coherence of the application of our concepts to possible worlds already involves us in thinking about entities, things, most literally understood. Thus, reflections of the "practical kind" are not, in virtue of their practicality, non–metaphysical. Rather, thinking about practices and their possibilities can be reveal ontology.

Think of a contrast here between the starting assumptions of Olson and Baker. According to Olson, the discussion of things must be based on some rudimentary principles of composition, some definition of what counts as a 'substance concept', and some other principles. This is why various relations that arise because of the contingencies of our social relations—like any normative concerns—are to be handled by ethicists. But, as we have seen, and as Olson is happy to admit, the claim that we are most fundamentally human organisms does not have to reflect our own self-understanding as persons. Or, as we learn from Baker (2008), this may just mean that we should not bother with this kind of metaphysics but assume that our ontology should reflect what we find significant and valuable in the world. I don't need to choose between the two here, but taking the second option is a respectable way to go: we allow that what we take to be significant and important, what we take to be dependent on our deep interests in our lives can inform our ontology. But then words like 'responsibility', 'anticipation', 'person' do not have to be seen as labels we attach to reality pre-cut at its properly metaphysical joints. In a butcher's shop of reality, the cuts we make—seemingly artificial from some purely abstract standpoint—are contingent on our nature, yet not arbitrary. In the world in which all of the relations that hold our practical lives together are compromised we should expect fundamental differences in psychology, and thereby in biology as well. What such changes might be may be an empirical question. The philosophical point is that if we are starting with some intuitive idea of 'person', given some particular background of intelligibility, and if our metaphysics is responsive to various parameters of such a concept, then spec-

ulations about where such a concept might break down are also speculations about the material conditions of survival, continuation, and so on, and vice versa.

Tellingly, as I discussed, some contextualizing along the lines that I am suggesting is obviously already present in various discussions of the social consequences of thought–experimental transformations, even though their presence in the discussion is usually secondary. The difference I suggest is that to call these considerations secondary, minor, or insignificant is to make a choice about what one takes to be relevant in addressing the question. As I argued earlier, once Parfit turns to the question of "what matters", we cannot occupy some neutral theoretical ground. Instead, we have to consider his question as addressed to us, and we are supposed to think about what matters, or what ought to matter, in survival, from where we are.[25] At the same time, this choice results in a tension between the theoretical fantastic tweaking that is offered for our contemplation and the background that we presume to be in place in order for what is proposed to be in any way intelligible. Or so I argued when revisiting the fission case earlier in this chapter.

For example, it is one thing to assert, on theoretical grounds, that "it is relation R that matters", but it is quite another to reflect on what that kind of assertion really means for a temporal entity like a person. What kinds of details one cares to reflect on show up in how much weight one might assign to thinking about further details, longer time-span, or the overall social environment. (And these choices are by no means neutral.) As I discussed earlier, Parfit does consider the consequences of his theoretical conclusions. For instance, Parfit tries to do it in giving us a contrast between being rationally convinced that there is no "further fact" of identity and actually maintaining the belief over time.[26] It is telling, though, that only for a brief moment—perhaps familiar to us, if only for the fact that the feeling is very fleeting—can he sustain this kind of conviction as a result of intellectual meditation. It is no surprise, then, that he finds Buddhism's detachment

[25] As Eileen John (2003) notes, discussing the value of detailed and "individualizing" thought (present in literary fictions and typically absent from the schematic presentations of thought experiments), such contextualizing by elaborating and reflecting on further details shows what we think about the question we are addressing: the constraints and emphases show the space of possibility for how the scenario might develop (153).

[26] See "Is the true view believable?"

conceptually congenial, although he bypasses Buddhist practice's very challenging set of routines—the extraordinary *labor* that goes into achieving the enlightenment afforded by detachment from worldly attachment and assumptions. But the reasons Parfit's assertion is not convincing have everything to do with the intersection of psychology, embodiment, and world as mutually implicated in systematic interactions. Thought experiments, according to the view I am developing, need to put all these elements together in order to address the question of personal identity.

6.7 Conclusion

In this chapter, I discussed the "divide and conquer" objection to the methodological proposal outlined in the previous chapters. The objection is based on the idea that metaphysical investigations are not constrained by the investigations of value and practical concerns more broadly understood. Replying to this objection, I appealed to Marya Schechtman's discussion of the difference between direct questions of value and the question about the conditions of asking such direct questions of value. Based on this distinction, we can reinterpret Locke's question of personal identity as being about the proper unit of our concern, about which the more direct questions can be asked. Simply based on this distinction, there is logical space for thinking that not all discussions of practical concerns are confusing metaphysics with ethics. If practical concerns are used as a guide for proper understanding of the ontological status of persons, then opening up thought experiments for the incorporation of further details about the practical and value background of our lives is a welcome addition to the philosophy of personal identity. Of course, this is not the end of the story. Since practice is multifaceted, there is a great variety of our concerns out there. While as we just established, they may have a point of connection with metaphysics of identity, it is not clear why that connection isn't splintered precisely to reflect the multiplicity. I next turn to this objection.

Chapter 7

Plurality of Practical Concerns

7.1 Introduction

Let's remind ourselves where we are. In Chapters Two and Three, I looked at the role of fantastic scenarios in literary fictions and in bioethics to generate insights for the 'literary model of thought experiments' in the philosophy of personal identity. In Chapter Four, I applied the model to the famous cases of fission and teletransportation. By considering a fuller vision of the thought-experimental background, we get a more holistic understanding of the complex interrelations of different aspects of our lives. By putting the fantastic change in dynamic interaction with the background—by letting them inform each other as we rely on some shared understanding of the more and less likely possibilities of their development—we explore the contours of our form of life. Such imaginative engagement proceeds by stipulating the features of the world that are more or less likely to support the fantastic transformations pictured in thought experiments: we imagine the space of interpersonal interaction in the context of the possible world's institutions and practices and think about the shape of practical concerns in that world. Notably, in engaging with thought experiments in this way, we do not directly ask whether some person survives teletransportation or fission; rather, we explore the conditions in which such direct questions can be asked.

Tying answers to metaphysical questions to speculations about practical concerns faces

a predictable (and familiar) challenge, namely the objection that our practical concerns are not a good guide to metaphysical truths. I discussed this objection in Chapter Six, and argued that there is logical space for a metaphysical question of identity that is not independent of our practical concerns. I claimed that my model of thought experiments may be particularly suitable for the explorations connected to answering this metaphysical question.

There is another challenge, however, again targeting the central assumption of the mainstream discussion that there is an intimate connection between the metaphysics of identity and our practical concerns. While Olson challenged the psychological continuity theory as confusing the metaphysical with the practical, in several recent papers David Shoemaker has been articulating the pluralist picture of the relation between the metaphysics of identity and practical concerns. According to this picture, different practical concerns may be grounded by different metaphysical relations, psychological and biological, and searching for their "grounding unifier" is futile and may prevent us from appreciating the amazing plurality of our concerns, their groundings, and the possible relation between different clusters of such concerns. Since my arguments in the earlier chapters suggest that some unified entity is a structural presupposition of our thinking about ourselves, pluralism presents an objection to my model of thought experiments.

In this chapter, I want to spell out Shoemaker's proposal, and explore agreements and differences, or possible affinities, between his views and the methodological approach I am defending. It may seem that my approach to thought experiments is particularly congenial to Shoemaker's thought. By advocating further exploration of thought-experimental details of the background world, I may seem to be open to the kind of pluralistic approach he defends. This, however, is not the full story, and there are some tensions between pluralism and my model of thought experiments.

7.2 Pluralism

Let's look at the details of Shoemaker's proposal. Return for a moment to Olson. As you recall, Olson's project was to sever the metaphysics of personal identity from practical concerns. Part of Olson's strategy was to explain that the intuitions we typically tie to identity can be tracking a different, practical relation that does not have the formal structure of identity. Olson appealed to Parfit's argument that "identity is not what matters" to show that the defenders of the psychological approach should not at all be surprised since it is actually their view as well. As Shoemaker observes, however, there are two different senses of Parfit's "identity is not what matters". The weaker reading is Parfit's solution: we can have what matters without uniqueness, since psychological continuity was the primary ingredient in identity. On the stronger reading, however, neither uniqueness nor psychological continuity, as the ingredients of personal identity, matter (Shoemaker 2011a, 23). As Shoemaker notices, simply by exposing the consequences of Parfit's thesis, Olson has given a defense of the weaker reading. This does not divorce identity from practical concerns, however, because, according to Shoemaker, if psychological continuity is the primary constituent of identity, the proponent of the psychological approach can claim that practical concerns are connected to identity in this roundabout way. To sever the ties between identity and practical concerns completely, we need to go for the stronger reading, as he proceeds to do (Shoemaker 2011a, 24).

Shoemaker (2007) also acknowledges his debts to Schechtman's (1996) distinction between the reidentification and the characterization questions of identity. She argued that our practical concerns are properly tied to the answer we give to the second question.[1] But, again, Shoemaker thinks once we separate different *kinds* of questions of identity, we can consider whether each practical concern can be grounded by different metaphysical

[1]We are already familiar with the reidentification question from Chapter Five. The characterization question asks what makes an action, experience, and so on, attributable to a particular person. Thus, the main difference between these two central questions of personal identity, as I mentioned earlier, has to do with the difference between the relata that go into answering them. The reidentification question is about the relation between different persons or person-stages, while the characterization question is about the relation between a person and experiences, actions, and so on.

relations.

Shoemaker's pluralist account urges that attention to the particulars of practical concerns reveals them to have a plurality of grounds, (as opposed their being grounded by a single relation). According to Shoemaker, there are at least nine different practices and concerns that figure in our starting views about the connection between personal identity and practical concerns. They are: anticipation of the experiences of my future self, special concern for the person who is myself, survival, moral responsibility for the things I did or will do, compensation for my sacrifices, sentiments like embarrassment, pride and regret that I feel for myself, third-person reidentification, and first-person reidentification (Shoemaker 2007, 317–318). In what follows, it is sufficient to discuss a shorter list, which Shoemaker shares with Schechtman's (1996) "four features": responsibility, prudential concern, compensation, and survival. Shoemaker problematizes the assumption that there is one criterion of identity that can ground each of these practical concerns by appealing to several examples that make it apparent that what justifies some of these concerns is neither identity nor psychological continuity as it is usually portrayed.

Let's start with responsibility. Our starting assumption was that responsibility presupposes identity. But, first, as Schechtman (1996) argued, the relata of the relation of responsibility and the relation of identity are different. In fission, as Shoemaker observes, we are presented with precisely the conceptual wedge between the two. Neither of my fission products is identical to me, but both seem to bear the appropriate relation with respect to my actions in the past. That is, ownership does not presuppose identity (Shoemaker 2011a, 28).[2]

Moreover, identity is one-one, while attributability of an act is not, which is clear from actual, as well as fantastic, cases. For example, if we sing a duet, the act itself is attributable to both of us. If we assume that attributability presupposes identity, what makes me responsible for singing the duet must be the holding of the relation of identity

[2]Shoemaker rightly gives credit to Schechtman 1996 for distinguishing between reidentification and characterization questions of identity. Ownership question is just characterization question, but without commitments to narrative identity.

between me and the singer. But the same goes for you. So, we are identical to each other. By reductio, we should deny the assumption that attributability presupposes identity (Shoemaker 2012, 24).[3] It gets even worse. A general may be held responsible for "taking the bridge" if it was taken on his orders, despite the fact that he was nowhere near the bridge, has never seen it, etc. (Shoemaker 2012, 26). In such chain-of-command actions, ownership can be attributed to a person without the person's actual performance of the deed, as Shoemaker argues. So, clearly, responsibility does not presuppose numerical identity.

Now, one might think that ownership with respect to responsibility still presupposes psychological continuity, as it is described by the proponents of the psychological continuity theory. But, as Shoemaker (2011a) argues, it is not psychological continuity that grounds this kind of ownership, but more likely only a subset of psychological relations—the "preservation of the psychological elements contributing to one's volitional network," since the psychological continuity theory is too broad of a relation (29).

Now consider anticipation. Again, it looks as if anticipation presupposes identity. But in both fission and fusion, according to Shoemaker, the pre-fission and pre-fusion person(s) can anticipate the experiences of her descendant(s) without identity holding after the procedure. What is required, according to Shoemaker (2011a), is only the "properly connected stream of conscious awareness" (30). Notice that this stream of conscious awareness is not the same as the one that grounds responsibility since one can be aware without being responsible; nor is it, again, exactly psychological continuity as described by the psychological continuity theorists because it is only a portion of it.

Now turn to self-concern. It can be argued that ownership with respect to it requires a "kind of persistence or resemblance relation between the values and attitudes directed towards the relevant object (the entities filling the roles of self, friends, family, etc.), along with connections of memory (or q-memory), intentions, beliefs, desires, and goals relevant to those cared-for objects", which, as Shoemaker admits, does begin to sound like full-

[3]In this thesis, the page references to Shoemaker 2012, "Responsibility without identity," refer to the paper that David Shoemaker kindly shared with me prior to the paper's appeared in print.

blown psychological continuity. But even if this particular practical concern presupposes psychological continuity, it does not presuppose identity, as is demonstrated by duplication (Shoemaker 2011a, 31).

Finally, for ownership with respect to compensation, all that matters is that it is "attached to physical/psychological self, and this will be determined by the specific nature of the benefit and burden in question" (Shoemaker 2011a, 36). Identity is not necessary since one can be compensated for playing on a team without being identical to the whole team, as we saw earlier. Moreover, psychological continuity is not necessary either. As Shoemaker argues, if physical damage occurs, one can be compensated as a biological organism whatever the psychological continuity or discontinuity between psychological selves. The idea here is that even if I undergo a profound psychological change, there are still grounds for compensating this new person for previously suffered biological damage. The grounds for such compensation are generated by the continuity of the biological organism (Shoemaker 2011a, 33; Shoemaker 2007, 338).

In each of the cases we discussed (and there are, of course, more practical concerns to turn to, but these four suffice to make the point), the relation that grounds each specific practical concern is different. What grounds responsibility is not what grounds compensation; what grounds anticipation is not what grounds responsibility; and so on. In sum, different kinds of psychological relations we form make us different kinds of owners, whereas we can be identical only to ourselves (Shoemaker 2011a, 39).

But isn't it clear that all of these concerns are *ours*, and that that is the importance of identity? According to Shoemaker (2011a), this proposal is too general to be of much help: "if the role they [criteria of identity] would play in justifying attributions of responsibility [for example] failed to illuminate those attributions in any real way, they would fail to be relevant to the practical concern in the way many have thought identity to be" (8). Recall that according to our starting assumption, a criterion of identity was supposed to provide the justification or proper account of ownership of actions and experiences that figured into our understanding of practical concerns. But the details of Shoemaker's argument

compromise that simplistic assumption. In each case, as we see, identity does not play the informative justifying role in explaining the details of holding somebody responsible or choosing to compensate somebody for this or that.

One can see that getting the picture about these details right may have important implications for clarity about our practices and their foundations. Even more importantly for our context, identifying the mistake at the outset forestalls many of the intractable difficulties generated by it. Shoemaker (2011a) concludes that identity is "the reddest of herrings": what does the explanatory work is the specific ownership relation in each particular instance of practical concerns (39).

As I said, Shoemaker admits that his project is very much inspired by Schechtman's (1996) work on the characterization question of identity, but without the assumption that all of the questions that arise in the contexts of our practical concerns are unified in virtue of narrative identity. Indeed, for him, it is an open question whether they can be brought together in virtue of any underlying unity at all. The main lesson of pluralism, as I am reading it, is to question the assumption that there is any such unity underlying all of our practical concerns. But, as Shoemaker (2007) admits, this lesson opens up new avenues for investigating whether and how different practical concerns and their grounding relations may interact.

7.3 Motivating Responding to Pluralism

Shoemaker (2011a) argued that the ownership relations with respect to different practical concerns are distinct and "thwart genuine unity in anything but name" (41). According to the general pluralist picture, we have different concerns distributed over our animality, humanity, personhood and what not. Forcing a unity on them misses the pluralities of how our concerns may be grounded in different aspects of our being. While Shoemaker (2011a) does not deny that there may be a question about the *locus*,[4] about which all of the more specific questions can be asked. However, he argues that that question is

[4]I discuss Marya Schechtman's work on identifying such locus of concern later in the chapter.

not going to help us with justification, since the question of locus is too general to tell us something informative and explanatory about practical concerns. (More on this in a moment.)

Let's consider an example to make our discussion more concrete. Suppose that after a terrible accident we have to perform a brain transplant on a relative. The outcome of the operation is that the relative's body with some parts of the brain is kept biologically alive with some severely limited functioning, while the receiver of the transplant inherits the psychological profile of the original: memories, beliefs, projects, etc. Suppose, in addition, that the relatives and friends are torn between their allegiance to the psychological continuer, on the one hand, and the person who remains alive and even somewhat functional, but loses the psychological traces of the original relative, on the other.[5] Suppose, then, that the mother visits her unfortunate child who does not remember her, while the sister and friends begin to enjoy the company of the person who remembers and loves them.[6] (Shoemaker claims that parental concerns, for example, track our biological nature [Shoemaker 2011a, 26].)

The case is no doubt complicated and prompts a lot of questions, some of them more general than others. The more general questions are: Whom is each visiting? and What happened to the person who underwent the procedure? The more particular are: Whom should the authorities compensate, for example, for the sacrifices that were made before the accident? (Who inherits the pension funds necessary to support decent living conditions, for example?) Who is now responsible for keeping the promises made in the past? And so on.

We have several options here. The main lesson of pluralism seems to be that our practical concerns are not unified in virtue of being grounded by the relations of numerical identity or psychological continuity. Instead, we are supposed to inquire into the nature

[5] Olson's (1997) version of Locke's Prince and the Cobbler is my model here (42).

[6] Of course, both of the survivors suffer changes in both biology and psychology, so the description in the example is crude. But all I need for the case here is that somebody with a superficially familiar body but with a dramatically different mental life is taken by the mother to be her child, while the sister and the friends take the person who remembers them and loves them, whatever happened to the old body, to be their loved relative and friend.

of each ownership relation directly. Let's first think of ownership with respect to the past burdens of pension payments. Which of the survivors—biological or psychological—should now be compensated for those burdens? Can both own the burdens of the past? In this case, it may be that both can be the owners, and that the funds should be divided based on some additional considerations of the other details of the case. For example, if the job that earned pension payments resulted in burdens more or less on the biological level, it is the biological organism that now needs treatment that may have to be compensated, and so on.[7] At the same time, of course, we may think that it is the psychological continuer who in fact remembers the burdens and looked forward to being compensated. I am not saying that there is some clear–cut metric we could use to determine the precise manner of distributing the compensation, but one can attempt to argue this way. Notice that neither of the survivors will be happy with the arrangement, but they may just have to live with half the funds.

What about other concerns? Who should keep the promises that the original made (i.e., what about the ownership with respect to responsibility)? It seems that it must be the psychological survivor because she has inherited the memories of making such promises, identifies with the agent in the past who made those promises, and in general has the projects and cares of the original. So, it looks like the person who lost the original's psychological profile does not stand in the relation of ownership with respect to the promises that were made, whereas the other survivor does. (These kinds of intuitions are typically used to support the psychologically-based accounts of personal identity, like Parfit's.) Maybe we can adopt the pluralist picture, then.

Things are not as simple, however, when it comes to reflecting on the more involved and complicated personal interactions that we encounter when we consider the more fully worked-out details of an on-going social life in a thought -experimental situation. Recall our discussion from Chapter Five. There, I insisted on the important roles of considering the broader contexts in which thought-experimental stories find their home and thinking

[7]David Shoemaker, in correspondence, prompted me to think of this option.

about the complicated interactions between the background of our form of life and the fantastic transformation of the individual. If we focus only on particular questions of compensation or responsibility as the psychological relations that underlie our practical concerns, we may lose track of the broader context in which inquiring about these relations makes sense in the first place.

I think that even though we may be able to resolve the issue of compensation for past burdens by dividing the money in some way, further considerations bring forth that the mother and the sister cannot so easily separate their third-person identification concern from the responsibility and compensation concerns. If we add more details into the lives of the survivors, we may see that it is more plausible to think that each maintains that the individual she is visiting *is* the person who suffered the accident and not just the vehicle for the specific grounding relations of separate practical concerns. That is, their disagreement is not simply at the level of resolving particular practical concerns about where compensation should go, or who is responsible. Instead, they can be seen as disagreeing about the question about which of the survivors continues the life of the original. This general question about the relation between the original person and the thought-experimental products is different from the particular questions about the practical concerns, although the first question is obviously related to the second. According to this interpretation, it is not simply that the mother is visiting her child qua biological survivor, and the sister enjoys the company of her sibling qua psychological survivor. Instead, they each think—given how our world is—they are spending time with *the* survivor.[8]

So the disagreement between the mother and the sister can be interpreted as being about what set of relations are to be given more weight, let's say, in tracking the individual, rather than only tracking the particular grounding relations for the particular practical concern. Whoever is right, and whether there can be a resolution of such a difficult case,

[8]David Shoemaker pressed me to clarify this. Am I saying that this is what is going on on all tellings of the story? Can't there be a pluralist telling of the story? Am I not loading the bias into the additional details by preselecting the details which favor a particular telling of the story? This is a complicated issue. At least, it shows that there may be different tellings of the story, but then this shows that the issue is far from settled. My view does not claim that such stories are conclusive. Their value partly lies in further discussions that they can generate.

the compartmentalization solution may sound neat at the outset, but goes against how we think of people in our daily lives. The disagreement cannot be dissolved, then, by pointing out that each may be overgeneralizing their particular inclinations with respect to their favorite practical concern and by asking them to consider the pluralist solution. Or, if we think that in such rare and anomalous cases we in fact should legislate giving up on the intuitive unity of a person, we need more of an account of what further changes in our practices this would entail, to see if we do not in fact run into some inconsistencies.

To summarize, assuming that practical concerns, according to Shoemaker, "just aren't unified," the social coordination of practices with respect to our unfortunate individual may involve separating different practical concerns and trying to figure out whether the relations that ground the proper attribution of such concerns go to this or to that individual. Some types of ownership may go with separate individuals (e.g. responsibility), while others may not (e.g. compensation). As I suggested, however, reflection on the further details of the case prompts certain worries about the theoretical approach Shoemaker advocates. Because we have two individuals who have some claims to the social space which in normal contexts is occupied by a unique person whose psychological and biological features are tightly intertwined, the inevitable conflicts introduced by the competitor require further reflection on the usefulness and motivation for the suggested compartmentalization.

7.4 Locus of Practical Concerns

Before moving on, it will be worth spelling out, however briefly, what one might mean by the idea of a locus of our practical concerns. I will not be defending any particular understanding of the issue, and only bring this up to facilitate the discussion that follows. We saw in the previous chapter that Locke's approach was to tie our practical concerns to sameness of consciousness. According to this approach, 'person' is a normative concept that we track by tracking our practical concerns. We saw both Olson's criticism of this approach and Schechtman's attempt to salvage Locke's original insight from Neo-Lockean

developments by distinguishing between two levels of questions about personhood. Now Shoemaker argues that different practical concerns go with different grounding relations: both biological and psychological. Lockean sameness of consciousness cannot do justice to the various practical concerns that are associated with bits of our personal trajectories before and after acquiring the robust sameness of consciousness: when we were fetuses, for example. Is there some other way that all these concerns can be brought together?

In recent papers, Marya Schechtman (2008, 2010) has been articulating and defending the 'person-life' view (PLV) of our identity that avoids the pitfalls of both the psychological continuity theory (which is subject to the extreme claim and denies that we were fetuses or can become vegetables) and Olson's animalism (which does not seem to directly connect to practical concerns). According to her, we can appreciate the insights that both camps offer us by expanding the idea of practical concerns to incorporate our animalhood more directly than the psychological continuity theory does, while arguing that the biological essentialism of animalism, as championed by Olson, is an abstraction from what is a more accurate anthropological understanding of human beings,[9] for whom culture is not simply added on to our animalhood, but rather reconstitutes it in very important ways. Thus, we are practical animals (human animals) through and through.

According to the person-life view, "personal identity consists in the continuity of a person-life; a person persists as long as a single person-life does" (Schechtman 2008, 37). It is a "kind of life typically lived by an enculturated human" (Schechtman 2010, 278). The boundaries of a human life define the boundaries of a person. Within the context of a person-life, we can ask questions of personal responsibility, compensation, and other practical concerns. We should think of a person-life as a trajectory or a career that a typical human organism follows (Schechtman 2010, 279). The practical concerns that take on an overly psychological tone in the psychological continuity theories here are fused with and permeate all of our human functions. Our animal is thoroughly cultural, and there is little hope in separating the shape that our practical concerns take from our

[9] I borrow the term 'anthropological' from Shoemaker 2011a, 15.

animal nature; both are mutually constraining. One way to put it is to say that PLV does anthropological justice to animalism. For our purposes, Schechtman's PLV account is an account of the basic unit (locus), about which it is proper to raise the more specific questions of our practical concerns, or in the context of which the practical concerns are properly addressed.

How can this account address the challenges posed by Shoemaker? Shoemaker (2011a) says that if one were to use PLV to save the general project of attempting to unify all of our practical concerns, PLV gives a robust but uninformative answer because it obscures the distinctions between different concerns and also between different grounding relations (43). In what follows, I challenge Shoemaker's view that there is no informative role that the question of locus can play in the quest to understand the connection between practical concerns and metaphysics. That is, even if Shoemaker can show that there is disunity of our concerns *with respect to the particular details of their metaphysical grounding*, considering the general question about the nature of the owner/locus may play an informative role in our understanding of practical concerns. What motivates my resistance to thoroughgoing pluralism is very simple. It seems that in order to discuss ownership (of any kind), we need some fairly stable conception of the owner of the said relations and the owner's persistence conditions. And it seems that the nature of ownership depends on what kind of thing the owner is.

In fact, Shoemaker himself suggests the possibility of different *kinds* of grounding for metaphysical relations with respect to practical concerns. According to this proposal, metaphysical relations can play a different role with respect to practical concerns than that of grounding justifiers. Shoemaker argues that, while failing to justify our practical concerns, metaphysical relations of this or that kind can, all the same, make such concerns *possible* in the first place. He calls this kind of grounding "rendering possible" (Shoemaker 2007, 346–347). For example, the metaphysical fact that there exist relations of psychological continuity between me and my loved ones may provide an explanation for why in some cases my loved ones may be compensated for the burdens I have suffered

(Shoemaker 2007, 347). According to this line of reasoning, 'grounding' here may be "taken to refer to what provides sufficient conditions, in our case a metaphysical set of conditions rendering the practices in question possible" (Shoemaker 2007, 347).

Building on this suggestion to expand the repertoire of the functions that thinking about metaphysics can play in our understanding practical concerns (and, in parallel, expanding the range of options with respect to the relation between the two in general), I will argue that pluralism with respect to justification does not yet show that there is no informative role that considerations of identity (or some genuine unity)—of whatever kind—can play. We will first consider an argument for the practical necessity of unification, and then later an argument for the structural dependence of practical concerns on the assumption of a unified locus of such concerns.

7.5 Unity in Practice

It is no news that once we are discussing agents who find themselves dealing with complicated nuances of their lives, we observe significant interaction between different practical concerns. The practical integration of different practical concerns must influence both the content and the phenomenological attributes (or quality) of the psychological relations that go into those individual concerns. I will claim that theoretical teasing apart of self-standing practical concerns (like responsibility, compensation, anticipation, etc.) from the holistic web in which they are found in real-life situations may misrepresent the relation between them. It assumes that their co-occurrence in an individual life does not have an effect on the nature of each individual concern. Thus, the pluralist move of disentangling different practical concerns from one another—however helpful it may be for some practical and theoretical purposes—cannot then be presented as if the results are independent of the initial unity of all these concerns as applying to the same person. The identity conditions and intelligibility of individual relations of responsibility, compensation, anticipation, and so on, are dependent on the broader range of presuppositions about what the owners of such relations are.

To illustrate this motivating idea, consider the case of working long hours on some long-term project like writing a dissertation. Clearly, the work of the will that takes responsibility for this project is connected to various beliefs about the expected outcome, is sustained by different kinds of current and future compensations and rewards, requires complicated adjustments in other commitments, and so on.

According to pluralism, the particular psychological relations that ground ownership with respect to responsibility do not coincide with the particular grounding relations with respect to anticipation. To say that all of the practical concerns are *ours* does not add anything in terms of explaining the connection between practical concerns and metaphysics of persons—it is not specific enough. If we want to know whether some future person is responsible for finishing the dissertation, we need to look only for the similarity of the volitional elements between the person who started the dissertation and the one who finished it, and not for any other strands of psychological connections, according to Shoemaker. In strange cases, while one and the same person may be both responsible for and anticipate the completion of the dissertation at time t1, at time t2, one person may be responsible because the elements in her volitional web are similar to the person who started the dissertation, while possibly a different person may be able to anticipate the joy of completing the dissertation as long as there is a proper stream of the psychological relations that can secure awareness of the completion of the dissertation.

I think this approach can be questioned along the following lines. In a fuller description of the case, the bonds that form between the experiences one goes through when taking responsibility for this particular project involve expectations about the type of compensation one may or may not find appropriate for it, the feedback effect in which one's anticipation of the proper reward fuels and strengthens one's attachment to the project, the necessary adjustments to one's feelings about other projects, the physical regime one has to keep to meet the requirements, and so on. Our practical concerns in such contexts—including the ones seemingly less connected to the project itself, like maintaining physical health, emotional balance, etc.— are in constant dialogical flux and

are clearly interdependent. Thus, the relations of continuity of the subset of the agent's volitional web that are tied to ascriptions of responsibility are in continuous interaction with the stream of conscious awareness grounding anticipation, the appropriate future– and past–directed attitudes, the strength of other relations of psychological and physical continuity and connectedness, and so on. Suppose you have the will to write the dissertation. Talk about the will will remain an abstraction, however, without the supporting web of beliefs about the value of engaging with such projects and various ideas about what one can accomplish, or not, by writing "the damn thing," including the risks of failing. And can one really anticipate the joy of completing the dissertation without in fact having been the person who has the will to write it? If history of long-term interaction between the volitional elements that go into constituting that will and the accompanying beliefs about the compensation, the anticipation, and so on, is not there, then it is unclear whether we have the identity conditions in place for counting the present will as properly similar to—or connected to—the past will that decided to write the dissertation. I.e., if enough of the rest of the psychological and biological elements of one's life are missing from the overall context in which the will is embedded it is not clear how to judge the similarity of the current will and the past will with which the agent currently identifies.[10] Considerations of this kind convince me that the bonds between these strands of different relations *reshape* the particular grounding relations of particular practical concerns, considered separately in abstraction from their multifaceted interactions. Without the supporting elements of the holistic web of our psychological relations, it is unclear how we can track the individual psychological strands through time. But then what becomes of Shoemaker's notion of grounding by metaphysical relation with respect to individual practical concerns?

The project of unifying different practical concerns into a coherent, non-disjointed life can be taken to be one of the primary objects of our care, even if we may never explicitly formulate this or be aware of it. While pluralism may be helpful in identifying

[10]See Schechtman 1996 for a development of a narrative self-constitution view of personal identity that emphasizes such interconnections.

the differences between various practical concerns we have, there is a certain neglect of the holistic considerations in its procedures. My proposal is that the reshaping of various practical concerns by others and by the context of their interaction undermines the idea that these concerns, and their grounding relations, are easily separable. I claim that our concerns would not be what they are without the holistic background of their occurrence in our form of life. More generally, to identify a specific practical concern as that of responsibility, compensation, and so on, we have to presuppose certain elements of our form of life, one of which is that there is a unified locus of our concerns that is reliably tractable and identifiable over time. Or so I will argue in the next section.

One might wonder, though, whether in my wish for a metaphysics that is thoroughly practical I have gone too far to the practical side. What I am discussing, it will be said, has to do with the practical necessity of unification, whereas Shoemaker's concerns are metaphysical. This proposal should remind one of some of the issues raised in the work of Christine Korsgaard (1989, 1991), which I will now discuss.

Arguing against Parfit's view that there is no deep unity of self-consciousness, Korsgaard claims that different experiences may still be unified by the requirement of "leading one life" (Korsgaard 1989, 113). In very rough and bold strokes, Korsgaard's (1989) proposal is that unification, synchronically and diachronically, is imposed on us by the practical necessity of making choices between competing drives by occupying a deliberative standpoint (110–111). The kinds of desires, projects, and relationships we thus choose can by themselves carry us into the future since most such things are extended over time. Furthermore, such projects compromise the idea of a "merely present self" since thinking in thick human categories of projects, character, habitual action and so on, requires that what I think of myself now is already fraught with how I see myself in the future (Korsgaard 1989, 113–114).

Thus, Korsgaard argues that the deep unity is not metaphysical in nature, but stems from the practical necessity of leading any kind of life that is recognizably like the lives of persons. To act, you need to deliberate; to deliberate, you need to view yourself as a

practical agent with a particular life; and lives are unified.

However, it may seem that Korsgaard only shows that while *really* disunified (because of Parfit's reductionism), we need to live *as if* we are unified (because of practical life). I.e., it may seem that the proposed unification is *merely* practical, and so it cannot really speak to the metaphysical concerns. This reading is prompted by Korsgaard's own language of "regarding ourselves as" unified, "constructing an identity" for ourselves, and so on (Korsgaard 1989, 109, 112)[11] It is this feature of Korsgaard's work that is connected to the objection that the necessity of practical unification does not translate into the requirement of the metaphysical grounding unifier.

This reading, however, misses another element of Korsgaard's view, which is no less important. Korsgaard (1989) claims that practical necessities can be "overwhelming" (115). I think we can understand this claim by focusing on Korsgaard's (1989) view of consciousness as a feature of different kinds of activities that we participate in (118). Many things that we do have this kind of feature. But since we are practically required to integrate different kinds of activities, we have to integrate that feature of them as well: i.e., we have to integrate different conscious elements that are part of such activities. As Korsgaard (1989) puts it, "[t]he phenomenon of the unity of consciousness is nothing more than the *lack* of any perceived difficulty in the coordination of psychic functions" (119). But then the order of dependence goes from the practical requirement of unified action to the unity of consciousness: "the unity of consciousness is simply another instance of the unity of agency" (Korsgaard 1989, 119). So, it seems that unity of agency can be presented as prior to the idea of unity of consciousness.[12]

[11]Consider what Korsgaard says in *The Sources of Normativity* (1994): "The conception of one's identity in question here is not a theoretical one, a view about what as a matter of inescapable scientific fact you are. It is better understood as a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking. So I will call this a conception of your practical identity"(101). Here it seems that we have to adopt some description and apply it to something more literal, like the life of a human being.

[12]In more recent work, Korsgaard argues that the life of human beings is not simply the result of adding more capacities to some more primitive notion of biological life. Rather, because human beings can choose, "personhood is quite literally a form of life.... Since being a person, like being a living thing or an animal, is a form of life, being a person is being engaged in a specific form of the activity of self-constitution" (Korsgaard 2009, 128–129). The connection between cultural activities should sound

Thus, experiences and actions which we are supposed to anticipate and for which we are expected to take responsibility already presuppose practical identity. As Korsgaard (1989) suggests in a footnote, her way of thinking about unity may push us to recognize that even the *idea* of "momentary experience" is suspect (117). I.e., practical identity is already constitutive of every experience and action that I take to be subject to *my* deliberation in how to act. In sum, for Korsgaard agential (or practical) identity is as basic as it gets.[13] This kind of identity is what gives us new reactive attitudes and allows for the practices of responsibility, answerability, etc. (Korsgaard 2009, 129); i.e., responsibility, for Korsgaard, is not possible as a practice unless there is the form of identity that can sustain such a practice.[14]

I think this gives us the sense in which the practical necessity of unifying our on-going cares, concerns, relationships, etc., is what makes you a person metaphysically, and not just practically, speaking. If even the notion of experience presupposes practical identity, the idea that individual practical concerns do not presuppose such identity is suspect. While this idea is available as an abstraction, its availability nevertheless depends on the holistic picture of our life, in which our concerns are unified in virtue of being part of it.

To summarize, I have suggested that the shape of each individual practical concern depends on some presupposed notion of practical identity. Now, this kind of appeal to psychological holism may not be convincing. Despite what Korsgaard argues, there is no implication here from the holistic interconnections of practical concerns and the difficulties of teasing them apart to the idea that there is a structural requirement for the uniqueness of the locus of our practical concerns. In what follows, I provide a different way to connect the kind of psychological holism discussed here and the presupposition of

familiar from our earlier discussion of Schechtman, who admits the influence of Korsgaard's *Locke Lectures* (which were later published as Korsgaard 2009) on her thinking.

[13]The idea of just how basic Korsgaard takes agential identity to be can be gleaned from what Korsgaard says about it: "This kind of identity is in a deeper way the person's own than an animal's identity, because he is consciously involved in its construction. And it is more essentially individual than a non-human animal's because he is free" (Korsgaard 2009, 129).

[14]She says: "We are responsible for our actions not because they are our products but because they are us, because we are what we do" (Korsgaard 2009, 130). Korsgaard (1994) speaks of "inescapable pervasiveness of moral identity" (122, 123) and, in a different work, that "[y]ou are one continuing person because you have one life to lead" and not the other way around (Korsgaard 1989, 113).

a unique continuer—of some kind—that our form of life seems to contain as one of its structural features.

7.6 Plurality of Concerns, Unification, Thought Experiments

Building on what I said about the practical requirement of the unification of our practical concerns and the parallel suggestion that pluralism misrepresents the dependence of the nature of each individual concern on practical or agential identity, I need to show that our understanding of different practical concerns depends conceptually on the structural requirement of the locus of our concerns as a fundamental component of the intelligibility of our form of life.

To focus the discussion, let's consider Shoemaker's (2012) case of the fissioning villain in more detail. Suppose you think that to hold somebody responsible for some past action requires the relation of identity between the person who performed the action in the past and the current person. Now imagine a villain who fissions right after committing some crime. Are the fission products responsible for what the villain did? According to Shoemaker, we have a strong intuition that both of the fission products should be held (at least partially) responsible. To confirm this, simply consider that both really enjoy the thought that they are getting away with something awful simply by fissioning. To *not* hold each of them responsible on the basis of non-identity (because two cannot be identical to one) to the original criminal just seems wrong. So, our intuition is based on the fact that each of the products stands in the relation of ownership with respect to the past action, and this relation of ownership is sufficient to ground the attribution of responsibility to each of them.

There are two ways to assess our reactions to the case of the fissioning villain. Shoemaker asks us to think about responsibility directly, by considering the relation in which each of the fission-products stands to that action. Since each of the survivors is psychologically continuous with the original, each of the two remembers the crime, plans to get away with it, etc. I think that the plausibility of the intuitive judgment that both of the

fission products are guilty in the case presented by Shoemaker depends on *assuming* the full practical background of this world. Thus, assuming that both of the fission products remember the crime, have criminal dispositions, present a threat to justice, and so on, we attribute the crime to both of them, relying on the established conventions of judging responsibility and attributability of actions with respect to responsibility. So, we take each of the products on their own, and we judge them as two separate people guilty of the crime. In effect, it is useful here to think of our reactions to this case as about the extensions of our current practices that we are prepared to tolerate at the margins.[15]

Notice, however, that such stipulations must implicitly rely on the background of the overall intelligibility of the practice of responsibility that we have in the actual world; in the case as presented, we are addressing the question *a*s internal to our practices. The idea would be that not punishing the fissioning villain would be the wrong extension of our practices because of our overwhelming intuitions that both of the fission-products are guilty. However, we could also pose the *external* question: Is fission a possibility such that it threatens the intelligibility of our practices in the first place? Asking and answering *that* question takes us out of the narrow focus on the limited set of psychological relations that obtain between both of the fission products and the original (memory, anticipation, etc.), and opens our investigation to a range of other features of the social and cultural background in which this transformation presumably takes place. This broadening of concerns is important because it presents a more complicated picture of the interaction between particular practical concerns and the overall background of our form of life. I will argue that our intuitive agreement in the fissioning villain case (that both are guilty) is no indication that "identity just gets in the way" of particular grounding. Rather, I think that it is only because some notion of identity is always there, as it were, that we can share such intuitions prompted by answering the questions *internal* to the practice.[16] When we start considering this case in the broader context of our lives, we can appreciate

[15] See Mark Johnston 1987, 1992, 1997; and Canfield 1990.

[16] Not, I take it, numerical identity in Olson's sense or the psychological continuity of the Neo-Lockean accounts.

the structural significance of the unified locus of our concerns in understanding each of the individual practical concerns. If such a locus is in fact a *structural* feature of our form of life, then we can show that the pluralist approach is at least incomplete.[17]

(The suggestion is not that we always have to address the external question before we ponder the internal question. We may in some cases gain clarity by thinking of internal questions first, or about both questions at once. But the availability of the other option, and a clear discussion of the differences between the two, will often uncover problems with focusing just on one of the alternatives.)

The question of considering the place of psychological relations in the larger social and cultural context is not new, of course. It is a familiar enough idea that the psychological continuity theory of identity must avoid accidental ways of securing the required connections. Suppose I have been brainwashed and now identify with, remember, and anticipate the consequences of some past action not committed by me. According to the idea that responsibility is grounded by psychological continuity, I am a legitimate candidate for attributing this act to me because I identify with it, remember it, and enjoy the thought of getting away with the crime. But is this ownership sufficient for responsibility? No. The *causal* story strikes us as being inappropriate for the legitimate attribution.[18] What does it mean to say that we also need the right causal story for proper attributability, a condition Shoemaker himself adds to the list of requirements for proper identification with some past self (2012)? Well, my identification with the past action has to be authored by me, and may not be a result of some accident. However, this fundamental distinction between the legitimate ownership of an action and the accidental generation of action finds its natural home in the overall scheme of things as we know them: the conditions where a person has had enough time to live through enough experiences to be able to make the right choice, to develop the right dispositions, and to display her character through her choices, among other things. To make coherent the idea that identification with the past

[17]Again, it is not part of this project to supply the positive response to the question of what that relation is.

[18]Even though there may be some bullet-biting responses that I am in fact responsible (Shoemaker 2011).

in the case of brainwashing is different from legitimate cases of identification, we thus have to appeal to some notion of proper cause. Such considerations, in turn, depend on a broad spectrum of issues that have to do with the overall shape of one's life. What is missing in the case of brainwashing is the rest of the story, the rest of one's life with others. [19]

Our questions about the fissioning villain may be put this way: when we judge the fission products to be responsible for the past crime, in what way do our reactions presuppose some notion of a unified locus of our concerns? Would this practice based on a kind of volitional psychological capacity normally exercised in a particular environment; i.e., would ownership of this kind be maintained without presupposing some notion of the socially accessible and easily traceable identity of the owner? Would there be *ownership* of the kind we have for the purposes of responsibility?

These are the kinds of questions that the literary model of thought experiments explores. This model asks us to speculate about the significance of different elements of our lives and to consider their complicated interactions. Reflecting on the general conditions of our judgments of responsibility or compensation, I think we may change our assessment of the thought experiment.

Let's reflect on such further details of the thought-experimental background. We start with the intuition that both fission-products have to be punished. (In the previous chapter, I discussed a different fission case, and the discussion here is continuous with the one I gave earlier.) As I argued earlier, we will end up with two individuals competing for the social space designated for one, and the competition ultimately problematizes much of what we can confidently say about individual practical concerns like responsibility.

Suppose each of the criminals tries to pick up the thread of life that the original led.

[19]Discussing a similar objection by Susan Wolf to her own Real Self View, Shoemaker writes that "the RSV as it stands focuses exclusively on how the self is related to its actions, independently of how that self is or is not related to the world and the people around it... What's missing from the RSV, in other words, is the condition of *normative competence*" (Shoemaker 2011). This notion is developed in Benson 1987 and Watson 1996. The objection that I am raising seems roughly similar in its spirit: what is missing in the case of brainwashing is the rest of the story, the rest of one's life with others, such that we can use the resources of this story to say whether the person in question is or is not responsible.

They both return to their family, friends, and enemies. This would be no easy matter. For example, take promises. The promises to them, or made by them, now are thrown off by the presence of another person who is liable to keep them and to receive what is owed to the original. But then the other person will not receive what is due to him. Can a promise be inherited by more than one person? How do they coordinate who carries it out? Notice that even phrasing the question of inheriting a given promise can be seen as problematic. If we think that the promise can be revealing of who one is, for example, then carrying out the promise is essential to the person's self-understanding, and is not reducible to some instrumental function of fulfilling the obligations one has.[20] So should both carry out the promise? Well, each may feel that he needs to do so in order to express himself. But we also need to consider the other side, namely the world, in which the promises are kept. Suppose you promised to help me, but your fission competitor has already helped me. You cannot then fulfill the promise. In some cases this will be trivial, but in others the inability to keep the promise will be debilitating because your social identity depends not just on your ability to make promises, but also on carrying them out, getting feedback from others, sharing the world with them. So promising, for example, cannot be easily inherited by a team of two, insofar as it is the kind of promising that matters in the sense of being expressive of who one is.

Similarly, fission threatens many of the practical concerns and relations that involve family members and close friends. What happens to the family of the original person who fissions? Can you be married to two of the fission-products? Can you appreciate now having two people who are your father where before fission you only had one? And so on. (I developed an analysis of fission along these lines in Chapter 4.) Reflections on these issues compromise the idea that fission is just more of the same (but doubled). Instead, it shows significant constraints on the intelligibility of practical concerns that come from their being embedded in a set of interpersonal and institutional relations that presuppose one-to-one correspondence between the act and the owner of the act. Once the

[20]In this discussion, I am indebted to Stanley Cavell's discussion of ethics in *The Claim of Reason.*

second owner is introduced by some artificial means, significant changes in the background required by this disruption threaten our ability to say anything about the status of the fission-products.

Fine. But isn't it still true that both of the criminals—whatever promises they keep or don't, whatever their families decide about them—should be punished? What we say about promising and families does not have to match what we say about responsibility. The advantage of pluralism is its splintering the presumed monoliticity of their grounding, and its approaching each individual practical concern directly. I think, however, that even our practice of responsibility is structurally dependent on some presupposition of the uniqueness of the continuer of a given life.

As I am arguing, one of the troubles with bare and abstract thought experiments is insufficient reflection on the overall background of our lives that provides the backdrop of the intelligibility of our practices. This applies to the way, in which we picture the fission products' psychology. We tend to have a narrow—synchronic—view of what their inner lives may be like. Thus, in our scenario, they are happy to get away with the crime. If we take the longer (and contextually embedded) view, we may change our mind about responsibility. Part of what I have been stressing in discussing promising, for example, is the role that one's ability to plan, anticipate, carry out, and get feedback about a given action plays in the overall trajectory of one's life. When we hold individuals responsible for what they do, we presuppose a holistic relation between the person and the place of a given action in her life. Exploring post-fission life introduces complications about the possible self-understanding that each of the fission-products may have.

Consider one more time the idea of action ownership as expressive of who one is. That is, instead of focusing on an action as an individual episode in a sequence of such episodes that constitute one's life, let's focus on the expressive powers of action: one's long hours of work on her language skills as showing her dedication to learning, one's patience in bringing up children as expressing her love for them, one's endless efforts to write that one book as expressing one's belief that he's got some talents, and so on, along with the

effects that they have on the world, the possibilities they open up, and the relationships with others that they foster. Now suppose that in each of these cases there is another person who also claims to stand in the relationship of ownership with respect to each of these actions. Now, what notions of action ownership, or of taking responsibility for one's actions should we expect to find in such a world? It may be particularly effective to consider the impact of introducing a co-owner on a child's developmental trajectory. In this case, the natural cycle of action–feedback–correction may be completely thrown off by the introduction of the competitor. One can only imagine what kind of effects this may have on a child's psychology. We can speculate about the details here, but in general the foregoing considerations are meant to show that our notions of action ownership and of taking responsibility for one's actions are deeply embedded in the world, in which the natural presupposition of a unique owner of actions is the default (barring collective ownership and such, which I think is a different story).

Now return to one possible development of our fissioning-villain scenario. Because of the significant difficulties associated with picking up the thread of life of the original person, experienced by both of the fission products because of the competition, we should expect some degree of self-reflection and reassessment on the part of such beings. The villain's original plan was to get away with the crime and to reap the rewards, but on reflection the plan will not play out as it should have: what we see is hard to describe as Parfit's double success. In some cases, this prompts a crisis and a dissociation from the past. So suppose that some fission-product criminals undergo a profound transition and claim that they are not continuing the life of the original, because their status as the ordinary continuer of the original's life is compromised for the reasons just mentioned. In this case, it is not as clear whether we would so easily hold the "reformed" fission-product equally as responsible for the past. Now, it may look as if this speaks in favor of pluralism because it seems to show that the reasons to diversify the treatment of the fission products is not based on identity considerations, but rather on the idea of identifying with an action. However, notice that the break with the past is prompted by the failure of

being the unique continuer, by the inability to pick up the thread of the original's life. If non-uniqueness can cause profound changes of this kind, then there is reason to think that it is intimately tied to one's ability to identify with some past action; consequently, there are reasons to think that uniqueness is tied to ownership with respect to responsibility.

Of course, the rebuttal to my response should suggest that not all criminals will experience the crisis of identity, and some may even learn to embrace the new opportunities of living the life of crime associated with the possibility of splitting. (Recall the analogous response that Parfit gave to Wolf's assessment of the consequences of fission. [Chapter 4]) Recall that in our case both of the criminals are thoroughly corrupted by evil and do not dissociate from the original person who committed the crime. In this case, then, isn't our initial intuitive judgment that both of the fission products are guilty accurate, indicating that there is no dependence on uniqueness in ownership with respect to responsibility?

This, however, does not immediately show that uniqueness is irrelevant. As I indicated earlier, in a fuller description of such a case we would have to investigate the difference in psychological dispositions of the original that allow the psychologies of the fission-products to cope with the complete overhaul of their lives. In this investigation, we may learn that in some (admittedly rare) circumstances, we will have exceptions to the rule. But the main idea I have been defending is that the requirement of uniqueness in our understanding of action ownership with respect to responsibility is structurally present in ways that are not detectable in the schematic description of the scenario that does not consider the contextual embeddedness of our actions in the broader context of the world. Even though speculations about the features of such worlds may not refute our starting intuition in the case as it has been presented, they may caution us about being too quick to rely on such cases.

The foregoing discussion shows, as I said earlier, the necessarily preliminary character of the conclusions reached by this way of thinking about thought experiments. For all we know, if fission became a real possibility, the shape of the changes in our practices and institutions might deviate from the possibilities we are now able to contemplate. In

fact, this is to be expected, and so the foregoing discussion cannot be seen as the final analysis. But this observation applies to any speculation about our practices, and so my view is on a par with other views that use thought experiments. In addition, it may seem that in order to have a conclusive response to pluralism I have to show that there cannot be possible worlds in which ownership with respect to responsibility does not presuppose some unified locus of our concerns.[21] I cannot show this, but as long as I succeed in showing that our entire vocabulary of the practice of ascribing responsibility depends, in some broad sense, on some stable unit for reidentification (not necessarily defined by the relation of numerical identity), I think it is enough to generate suspicion about accounts that imply that we can do without presupposing some such unifying locus of our concerns in thinking about practical concerns.[22]

To conclude, our intuitions in the fissioning villain case speak for holding both of the fission products at least partially responsible, but this judgment does not mean that 'identity just gets in the way'. The role of identity is apparent in further articulating the background against which we make more particular judgments. It is present in the shape of our practices being what they are.

7.7 Conclusion

In this chapter, I discussed David Shoemaker's pluralism, according to which different practical concerns are grounded by different metaphysical relations, and that the search for one relation that can ground them all—one of the central assumptions in the personal identity literature—is misguided. In response, I offered reasons to think that our practical concerns presuppose a unifying locus of our practical concerns, without which they would not be what they are. Such structural presupposition is uncovered by applying

[21] I am ignoring joint ownership and chain of command actions here since I think these are separate issues.

[22] There is a difficult question about the relation between what I am saying and Shoemaker's mention of *practical* identity (2011, 515). There, he says that this practical notion of identity can be defined in terms of practical identification, such that it will be possible for a person to be identified with others' actions. The proposal is to be further developed, and I hope that what I say may help in clarifying what the relation of practical identity may be like. But this aim is beyond the scope of the current project.

the literary model of thought experiments to the thought experiments used in arguing for pluralism. While the arguments may not be conclusive, they offer an alternative way of thinking about the complicated relations between metaphysics of identity and our practical concerns.

Works Cited

Appiah, K. A. (2008). *Experiments in ethics*. Harvard University Press.

Arras, J. (1993). Principles and particularity: the roles of cases in bioethics. *Ind. LJ*, *69*, 983.

Arras, J. D. (1991). Getting down to cases: The revival of casuistry in bioethics. *Journal of Medicine and Philosophy*, *16*(1).

Baker, L. R. (2000). *Persons and bodies: A constitution view*. Cambridge University Press.

Baker, L. R. (2008). Big-tent metaphysics. *Abstracta*, *2*, 8–15.

Baz, A. (2012). *When words are called for: A defense of ordinary language philosophy*. Harvard University Press.

Beauchamp, T., & Childress, J. (2008). *Principles of biomedical ethics*. Oxford University Press.

Beck, S. (2008). Going narrative: Schechtman and the russians. *South African Journal of Philosophy*, *27*(2), 69–79.

Bokulich, A. (2001). Rethinking thought experiments. *Perspectives on Science*, *9*(3), 285-307.

Brock, D. W. (1987). Truth or consequences: The role of philosophers in policy-making. *Ethics*, *97*(4), pp. 786-791. Retrieved from `http://www.jstor.org/stable/2381207`

Brown, J. (2008). Thought Experiments. In *Stanford encyclopedia of philosophy*.

Brown, J. R. (1986). Thought Experiments since the Scientific Revolution. *International Studies in the Philosophy of Science*, *1*(1).

Brown, J. R. (1991). *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. Routledge.

Brown, J. R. (2004). Peeking Into Plato's Heaven. *Philosophy of Science*, *71*(5).

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, *18*(1), 1–21. Retrieved from `http://www.jstor.org/stable/1343711`

Buchanan, A., Brock, D., Daniels, N., & Wickler, D. (2001). *From chance to choice: Genetics and justice*. Cambridge University Press.

Camp, E. (2009). Two varieties of literary imagination: Metaphor, fiction, and thought experiments. *Midwest Studies in Philosophy*, *33*(1), 107–130.

Canfield, J. V. (1990). *The looking-glass self: An examination of self-awareness*. Praeger.

Carroll, N. (2002). The wheel of virtue: Art, literature, and moral knowledge. *Journal of Aesthetics and Art Criticism*, *60*(1), 3–26.

Cavell, S. (1969). *Must we mean what we say?: A book of essays*. Cambridge University Press.

Cavell, S. (1979). *The claim of reason: Wittgenstein, skepticism, morality, and tragedy*. Oxford University Press.

Chiong, W. (2005). Brain death without definitions. *Hastings Center Report*, *35*(6), 20–30.

Clark, S. R. (1977). *The moral status of animals*. Oxford University Press.

Cummins, R. (1998). Reflections on reflective equillibrium. In M. DePaul & W. Ramsey (Eds.), *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry* (p. 113-127). Rowman and Littlefield Publishers.

Dancy, J. (1985). The role of imaginary cases in ethics. *Pacific Philosophical Quarterly*, *66 (January-April)*, 141-153.

Dancy, J. (Ed.). (1997). *Reading parfit*. Blackwell Publishers.

Danto, A. C. (1984). Philosophy as/and/of literature. *Proceedings and Addresses of the American Philosophical Association*, *58*(1), 5–20.

Davies, D., & Matheson, C. (Eds.). (2008). *Contemporary readings in the philosophy of literature: An analytic approach*. Broadview Press.

DeGrazia, D. (2005). *Human identity and bioethics*. Cambridge University Press.

DePaul, M., & Ramsey, W. (Eds.). (2002). *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry*. Rowan and Littlefield, New York.

Diamond, C. (1982). Anything but argument? *Philosophical Investigations*, *5*(1), 23–41.

Diamond, C. (1991). *The realistic spirit: Wittgenstein, philosophy, and the mind*. MIT Press.

Diamond, C. (2002). What if x isn't the number of sheep? wittgenstein and thought-experiments in ethics. *Philosophical Papers*, *31*(3), 227-251.

Eldridge, R. (2009). *The oxford handbook of philosophy and literature*. Oxford University Press.

Fodor, J. (1964). On knowing what we would say. *The Philosophical Review*, *73*(2), 198-212.

Gendler, T. (2002). Personal identity and thought experiments. *Philosophical Quarterly*, *52/206*, 34-54.

Gendler, T. S. (2000). *Thought experiment: On the powers and limits of imaginary cases.* Garland Pub., NY.

Gendler, T. S. (2004). Thought experiments rethought and reperceived. *Philosophy of Science*, *71*(5), 1152–1163.

Gendler, T. S., & Hawthorne, J. (Eds.). (2002). *Conceivability and possibility.* Claredon/Oxford University Press.

Gibson, J. (2007). *Fiction and the weave of life.* Oxford University Press.

Gibson, J. (2009). Literature and knowledge. In R. Eldridge (Ed.), *Oxford handbook of philosophy and literature.*

Glover, J. (1984). *What sort of people should there be?* Penguin Books.

Haaggqvist, S. (1996). *Thought experiments in philosophy.* Stockholm: Almqvist and Wiksell International.

Horowitz, T., & Massey, G. (Eds.). (1991). *Thought experiments in science and philosophy.* Rowan and Littlefield, Savage, MD.

Humphreys, P. (1993). Seven theses on thought experiments. In J. e. a. Earman (Ed.), *Philosophical problems of the internal and external world* (p. 205-227). University of Pittsburgh Press.

Jackson, F. (2001). Precis of from metaphysics to ethics. *Philosophy and Phenomenological Research*, *62*(3), 617-624.

John, E. (1998). Reading fiction and conceptual knowledge: Philosophical thought in literary context. *Journal of Aesthetics and Art Criticism*, *56*(4), 331–348.

John, E. (2003). Literary fiction and the philosophical value of detail. In M. Kieran & D. M. Lopes (Eds.), *Imagination, philosophy, and the arts* (p. 142-159). Routledge.

Johnston, M. (1987). Human beings. *Journal of Philosophy*, *84*, 59-83.

Johnston, M. (1989). Fission and the facts. *Philosophical Perspectives*, *3*, 369–97.

Johnston, M. (1992). Reasons and reductionism. *Philosophical Review*, *3*(3), 589–618.

Johnston, M. (1997). Human concerns without superlative selves. In J. Dancy (Ed.), *Reading parfit* (p. 149-180). Blackwell Publisher.

Jonsen, A., & Toulmin, S. (1990). *The abuse of casuistry.* University of California Press.

Kafka, F. (1988). *The metamorphosis, the penal colony, and other stories* (Willa & E. Muir, Trans.). Schoken Books, NY.

Kieran, M., & Lopes, D. M. (Eds.). (2003). *Imagination, philosophy, and the arts.* Routledge.

Kivy, P. (1997). *Philosophies of arts: An essay in differences.* Cambridge University Press.

Knobe, J., & Nichols, S. (Eds.). (2008). *Experimental philosophy*. Oxford University Press.

Korsgaard, C. (1989). Personal identity and the unity of agency: A kantian response to parfit. *Philosophy and Public Affairs*, *18*, 101-132.

Korsgaard, C. M. (1996). *The sources of normativity*. Cambridge University Press.

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

Kuhn, T. (1977). A function for thought experiments. In *The essential tension* (p. 240-265). University of Chicago Press.

Lamarque, P., & Olsen, S. H. (1994). *Truth, fiction, and literature: A philosophical perspective*. Oxford University Press.

Lamarque, P., & Olsen, S. H. (2004). *Aesthetics and the philosophy of art: The analytic tradition: An anthology*. Blackwell Pub.

Laymon, R. (1989). Cartwright and the Lying Laws of Physics. *Journal of Philosophy*, *86*(7), 353-372.

Laymon, R. (1995). Experimentation and the Legitimacy of Idealization. *Philosophical Studies*, *77*(2-3).

Lewis, D. (1976). Survival and identity. In A. Rorty (Ed.), *The identities of persons* (p. 17-41). University of California Press.

Locke, J. (1975). Of identity and diversity. In J. Perry (Ed.), *Personal identity* (p. 33-53). University of California Press.

MacIntyre, A. (1981). *After virtue*. University of Notre Dame Press.

Macklin, R. (2006). The new conservatives in bioethics: Who are they and what do they seek? *The Hastings Center Report*, *36*(1), pp. 34-43. Retrieved from `http://www.jstor.org/stable/3528596`

Mann, T. (1959). *The transposed heads: A legend of india*. Vintage.

Martin, R. (1998). *Self-concern: An experimental approach to what matters in survival*. Cambridge: Cambridge University Press.

Massey, G. (1991). Backdoor analyticity. In T. Horowitz & G. Massey (Eds.), *Thought experiments in science and philosophy* (p. 285-296). Rowan and Littlefield, Savage, MD.

Morrison, M. (2005). Approximating the Real: The Role of Idealizations in Physical Theory. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, *86*(1), 145-172.

Mounce, H. O. (1980). Art and real life. *Philosophy*, *55*(212), pp. 183-192. Retrieved from `http://www.jstor.org/stable/3750582`

Mulhall, S. (2008). *The wounded animal: J.m. coetzee and the difficulty of reality in*

literature and philosophy. Princeton University Press.

Munson, R. (Ed.). (1999). *Intervention and reflection: Basic issues in medical ethics.* Wadsworth.

Murdoch, I. (1971). *The sovereignty of good.* Schocken Books.

Nelson, H. L. (2002). What child is this? *The Hastings Center Report*, *32*(6), 29-38.

Norton, J. (1996). Are thought experiments just what you always thought? *Canadian Journal of Philosophy*, *26*, 333-366.

Norton, J. (2004). On thought experiments: Is there more to the argument? *Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association, Philosophy of Science*, *71*, 1139-1151.

Nozick, R. (1977). *Anarchy, state, and utopial.* Basic Books.

Nussbaum, M. C. (1990). *Love's knowledge.* Oxford University Press.

Olson, E. T. (1997). *The human animal: Personal identity without psychology.* Oxford University Press.

Olson, E. T. (2007). *What are we?: A study in personal ontology.* Oxford University Press.

O'Neill, O. (1989). The power of example. In *Constructions of reason: Explorations of kant's practical philosophy* (p. 165-187). Cambridge University Press.

Parfit, D. (1986a). Comments. *Ethics*, *96*(4), 832–872.

Parfit, D. (1986b). *Reasons and persons.* Oxford University Press.

Parfit, D. A. (1995). The unimportance of identity. In *Identity* (pp. 13–45). Oxford University Press.

Perry, J. (1975). *Personal identity.* University of California Press.

Perry, J. (1976). The importance of being identical. In A. Rorty (Ed.), *The identities of persons* (p. 67-91). University of California Press.

Perry, J. (1978). *A dialogue on personal identity and immortality.* Hackett.

Pillow, K. (2009). Imagination. In R. Eldridege (Ed.), *The oxford handbook of philosophy and literature.* Oxford University Press.

Rawls, J. (1971). *A theory of justice.* The Belknap Press of Harvard University Press.

Rescher, N. (2005). *What if? thought experimentation in philosophy.* Transaction Publishers, New Brunswick, NJ.

Rorty, A. O. (1976). *The identities of persons.* University of California Press.

Rovane, C. (1997). *The bounds of agency: An essay in revisionary metaphysics.* Princeton University Press.

Rovane, C. A. (1990). Branching self-consciousness. *Philosophical Review*, *99*(3), 355–95.

Sandel, M. (2004). The case against perfection. *The Atlantic*, *4*.

Schechtman, M. (1996). *The constitution of selves*. Cornell University Press.

Schechtman, M. (2004). Personality and persistence: The many faces of personal survival. *American Philosophical Quarterly*, *41*(2), 87–106.

Schechtman, M. (2007). Stories, lives, and basic survival: A refinement and defense of the narrative view. *Philosophy: The Journal of the Royal Institute of Philosophy*, *60(supp)*, 155-178.

Schechtman, M. (2008). Staying alive: Personal continuation and a life worth living. In C. Mackenzie & K. Atkins (Eds.), *Practical identity and narrative agency* (p. 31-56). Routledge.

Schechtman, M. (2010). Personhood and the practical. *Theoretical Medicine and Bioethics*, *31*(4).

Shoemaker, D. (2008). *Personal identity and ethics: A brief introduction*. Broadview Press.

Shoemaker, D. (2010). The insignificance of personal identity for bioethics. *Bioethics*, *24*(9), 481–489.

Shoemaker, D. (2011a). Moral responsibility and the self. In S. Gallagher (Ed.), *The oxford handbook of the self* (p. 487-521). Oxford University Press.

Shoemaker, D. (2011b). The stony metaphysical heart of animalism. *unpublished manuscript*.

Shoemaker, D. (2012). Responsibility without identity. *Harvard Revieiw of Philosophy*.

Shoemaker, D. W. (1996). Theoretical persons and practical agents. *Philosophy and Public Affairs*, *25*, 318-332.

Shoemaker, D. W. (2007). Personal identity and practical concerns. *Mind*, *116*(462), 317–357.

Sirridge, M. J. (1975). Truth from fiction? *Philosophy and Phenomenological Research*, *35*(4), 453–471.

Snowdon, P. (1991). Personal identity and brain transplants. *The Royal Institute of Philosophy Supplements*, *29*, 109-126.

Sober, E. (2001). Genetic determinism. In B. et al. (Ed.), *From chance to choice.* Cambridge University Press.

Sorabji, R. (2006). *Self: Ancient and modern insights about individuality, life, and death*. University of Chicago Press.

Sorensen, R. A. (1992). *Thought experiments*. Oxford University Press.

Stich, S. (1988). Reflective equillibrium, analytic epistemology, and the proglem of

cognitive diversity. *Synthese*, *74*(3), 391-413.

Stolnitz, J. (1992). On the cognitive triviality of art. *British Journal of Aesthetics*, *32*(3), 191–200.

Sullivan, I., C. W. (2001). Folklore and fantastic literature. *Western Folklore*, *60*(4), 279–296. Retrieved from `http://www.jstor.org/stable/1500409`

Thomson, J. J. (1999). A defense of abortion. In R. Munson (Ed.), *Intervention and reflection.* Wadsworth.

Walsh, A. (2011). A moderate defence of the use of thought experiments in applied ethics. *Ethical Theory and Moral Practice*, *14*(4), 467–481.

Wilkes, K. (1988). *Real people: Personal identity without thought experiments.* Oxford University Press.

Wilkes, K. V. (1981). Functionalism, psychology and the philosophy of mind. *Philosophical Topics*, *12*(1), 147–67.

Wilkes, K. V. (1986). Nemo psychologus nisi physiologus. *Inquiry*, *29*(June), 168–185.

Wilkes, K. V. (1991). The relationship between scientific psychology and common-sense psychology. *Synthese*, *89*(October), 15–39.

Williams, B. (1970). The self and the future. *Philosophical Review*, *79*, 161-180.

Williamson, T. (2007). *The philosophy of philosophy.* Blackwell Pub.

Wolf, S. (1986). Self-interest and interest in selves. *Ethics*, *96*(July), 704–20.

Wollheim, R. (1984). *The thread of life.* Yale University Press.

Yablo, S. (1993). Is conceivability a guide to possibility? *Philosophy and Phenomenological Research*, *53/1*, 1-42.

VITA

Aleks Zarnitsyn
Department of Philosophy, UIC
1430 University Hall, MC 267
601 S. Morgan Street
Chicago, IL 60607-7109 USA

Phone: +1 312 731 6087
E-mail: azarni2@uic.edu
Web: www.uic.edu/∼azarni2

Education

Ph.D. in Philosophy, University of Illinois at Chicago, 2013. Thesis Title: *Thought Experiments in Personal Identity: A Literary Model*. Primary Advisor: Professor Marya Schechtman

M.A. in Philosophy, University of Arkansas, 2004. Thesis Title: *Cavell and Rorty on Skepticism*. Primary Advisor: Professor Ed Minar

B.A. in Philosophy, with Honors, University of Arkansas, 2002

B.A. in German Language, University of Arkansas, 2002

Dimploma in Management, Cheboksary Faculty of Saint-Petersburg State Technical University, Cheboksary, Russia, 1999

Areas of Specialization

Personal Identity, Metaphysics (broadly construed)

Areas of Competence

Logic, Philosophy of Literature, Moral Philosophy, Bioethics

Presentations

"Can Fission Cut It: On the Cognitive Value of Thought Experiments in Personal Identity", Kentucky Philosophical Association Annual Meeting, Lexington, Kentucky, 03.23.13

"Fictioning Thought Experiments", American Philosophical Association, Central Meeting, New Orleans, 02.23.13

"The Cognitive Value of Fiction and Thought Experiments in Personal Identity", Institute for the Humanities, UIC, Chicago, 01.17.13

"The Cognitive Value of Fiction and Thought Experiments in Personal Identity" Fellow's Workshop, Institute for the Humanities, UIC, Chicago, 01.24.13

"Thought Experiments. Why Not?", UIC Undergraduate Philosophy Club, 10.01.12

"Organ Sales, Yuck Factor, Moral Theory", University of Minnesota at Rochester, 03.07.12

Awards, Fellowships, Grants

UIC Ph.D. Student Travel Award, February 2013

American Philosophical Association graduate student travel funding award. APA Central Meeting, February 2013

Institute for the Humanities Dissertation Fellowship, UIC (competitive university-wide fellowship). 2012–2013

Dean's Fellowship, UIC (competitive university-wide fellowship). 2011–2012

Institute for the Humanities Dissertation Fellowship, UIC. 2011–2012 (declined)

Ruth Marcus Graduate Student Award, UIC (departmental award). 2011–2012

Outstanding Teaching by a Graduate Student Award, UIC (departmental award). 2008–2009

Graduate Student Assistantship, Philosophy Department, UIC. 2005–2011

Philip S. Bashor Scholarship Award for outstanding graduate work in philosophy, U of A (departmental award). 2004

Graduate Student Assistantship, Philosophy Department, U of A. 2002–2005

Elizabeth Fullbright Study Abroad Scholarship, Karl Franzens Universitaet Graz, Austria. 2001–2002

Deutscher Akademischer Austausch Dienst (DAAD) German Academic Exchange Service Language Study Scholarship, Freie Universitaet Berlin, Germany. Summer 2001

Harold Hantz Scholarship for an outstanding performance as an undergraduate in philosophy, Philosophy Department, U of A. 2000–2001

International Student Scholarship, U of A. 1997–1998, 1999–2001

Teaching Experience

Courses Taught as Primary Instructor

Phil 102: Introductory Logic. (4 times)

Phil 210: Symbolic Logic.

Phil 105: Science and Philosophy (Evolution and Morality).

Phil 100: Introduction to Philosophy.

Phil 2003: Introduction to Philosophy (2 sections, 3 times).

Phil 2203: Logic—deductive and inductive reasoning (2 sections).

ILOG Introduction to Logic (critical thinking, informal reasoning, and symbolic logic) (3 times). Johns Hopkins University's Center for Talented Youth At Easton, PA

Courses Taught as Teaching Assistant

Phil 234: Philosophy of Film.

Phil 104: Introduction to Political Philosophy: Hobbes, Locke, Mill, Nozick, Rawls.

Phil 103: Introduction to Ethics: Stoics, Kant, Mill, Nietzsche.

Phil 100: Introduction to Philosophy: God, Free Will, Mind.

Phil 105: Science and Philosophy: Space and Time.

Phil 104: Introduction to Social and Political Philosophy: Citizenship and Immigration.

Phil 102: Introductory Logic.

Phil 2003: Introduction to Philosophy.

Phil 2003: Introduction to Philosophy.

Academic Service

Organizer: UIC Philosophy Graduate Student Workshop on long-term career preparation, 05.03.12

Session Chair: William Koch "Late Wittgenstein, the Mental, and Taylor", APA Central Meeting, Chicago, February 2009

Graduate Student Representative: Department Meetings, Fall 2008

Languages

German: (reading: advanced; writing: intermediate; speaking: advanced)
French: (reading: advanced; writing: beginner; speaking: beginner)
Russian: native speaker

Professional Affiliations

American Philosophical Association
Graduate Employees Organization, UIC

References

Marya Schechtman—Department of Philosophy, UIC
Email: marya@uic.edu; Office Phone: 1 (312) 413–7565

David Shoemaker—Philosophy Department, Tulane University
Email: dshoemak@tulane.edu; Office Phone: 1 (504) 892–3390

John Gibson—Department of Philosophy, University of Louisville
Email: john.gibson@louisville.edu; Office Phone: 1 (502) 852–0452

Colin Klein—Department of Philosophy, UIC
Email: cvklein@uic.edu; Office Phone: 1 (312) 413–1801

David Hilbert—Department of Philosophy, UIC
Email: hilbert@uic.edu; Office Phone: 1 (312) 996–5490

Walter Edelberg (Teaching Reference)—Department of Philosophy, UIC
Email: edelberg@uic.edu; Office Phone: 1 (312) 996–3022