

THOUGHT EXPERIMENT

ON THE POWERS AND LIMITS
OF IMAGINARY CASES

Tamar Szabó Gendler

GARLAND PUBLISHING, INC.
A MEMBER OF THE TAYLOR & FRANCIS GROUP
NEW YORK & LONDON/2000

Published in 2000 by

Garland Publishing, Inc.
A member of the Taylor & Francis Group
29 West 35th Street
New York, NY 10001

Copyright © 2000 by Tamar Szabó Gendler

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without written permission from the publishers.

10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Gendler, Tamar.

Thought experiment : on the powers and limits of imaginary cases / Tamar Szabó Gendler.

p. cm. — (Studies in philosophy)

Includes bibliographical references and index.

ISBN 0-8153-3656-X (alk. paper)

1. Thought experiments. I. Title. II. Studies in philosophy
(New York, N.Y.)

BD265 .G46 2000

128'.3—dc21

00-034074

Printed on acid-free, 250 year-life paper
Manufactured in the United States of America

Contents

| | | |
|---------|---|------|
| 0. | Preface | ix |
| 0.1 | Revisions Made | x |
| 0.2 | Summary of Contents | xi |
| 0.2.1 | Chapter 1—Introduction | xii |
| 0.2.2 | Chapter 2—Galileo (First Case Study) | xiii |
| 0.2.3 | Chapter 3—The Ship of Theseus (Second Case Study) | xiii |
| 0.2.4 | Chapter 4—Personal Identity (Third Case Study) | xiv |
| 0.2.5 | Chapter 5—Conclusion | xiv |
| 0.3 | Acknowledgments | xvi |
| 1. | Introduction | 1 |
| 1.1 | Exceptional Cases | 2 |
| 1.1.1 | Characterization of “Exceptional Case” | 2 |
| 1.1.1.1 | Exceptional Cases in Theories with Privileged Characteristics | 2 |
| 1.1.1.2 | Exceptional Cases in Theories without Privileged Characteristics | 6 |
| 1.1.2 | Ways of Accounting for Exceptional Cases | 8 |
| 1.1.3 | Patterns of Accounting | 9 |
| 1.1.4 | Application to the Question at Hand | 12 |

| | | |
|-------|--|----|
| 1.2 | Imaginary Cases | 14 |
| 1.3 | Thought Experiments | 18 |
| 1.3.1 | What is a Thought Experiment? | 18 |
| 1.3.2 | The Tripartite Structure of Thought Experiments | 21 |
| 1.3.3 | Three Sorts of Thought Experiments | 25 |
| 1.4 | Appendix to Chapter 1 | 28 |
| 2. | Galileo | 33 |
| 2.1 | Argumentative Reconstruction | 33 |
| 2.1.1 | The Elimination Thesis | 34 |
| 2.1.2 | Clarification of Terminology | 34 |
| 2.1.3 | The Negative Argument and the Positive Argument | 36 |
| 2.1.4 | The Dispensability Thesis and the Derivativity Thesis | 37 |
| 2.2 | Galileo's Thought Experiment and its Reconstruction | 39 |
| 2.2.1 | Galileo's Thought Experiment | 39 |
| 2.2.2 | Reconstruction of the Galileo Case | 41 |
| 2.2.3 | Four Ways out for the Aristotelian | 42 |
| 2.2.4 | What the Reconstruction Misses | 45 |
| 2.3 | Denying the Dispensability and Derivativity Theses | 47 |
| 2.3.1 | Rejecting Reconstruction: What the Thought Experiment Does | 47 |
| 2.3.2 | Rejecting the Positive Argument: What Makes these Beliefs New? | 49 |
| 2.3.3 | Rejecting the Negative Argument: What Makes these Beliefs Knowledge? | 53 |
| 2.3.4 | Constructivism and the Contrast with Norton and Brown | 56 |
| 2.4 | Conclusion | 62 |
| 3. | Theseus | 65 |
| 3.1 | Conceptual Thought Experiments | 65 |
| 3.2 | The Story | 67 |
| 3.3 | The Puzzle | 68 |
| 3.4 | Is the Ship of Theseus an Exceptional Case? | 70 |

| | | |
|---------|--|-----|
| 3.4.1 | Automatic and Specially-Secured Identity | 70 |
| 3.4.2 | Organisms, Artifacts, and Exceptional Cases | 72 |
| 3.5 | Attempts to Dissolve the Problem | 79 |
| 3.5.1 | Van Inwagen | 79 |
| 3.5.1.1 | Identity Under a Sortal Summary | 82 |
| 3.5.1.2 | Parfit | 84 |
| 3.5.2 | Attempts to Solve the Problem | 88 |
| 3.6 | A Traditional Solution: Hirsch | 88 |
| 3.6.1 | A Meta-solution: Nozick | 91 |
| 3.6.2 | The Proposed Diagnosis | 97 |
| 3.7 | Some Very General Candidate Principles | 99 |
| 3.7.1 | Remarks on the Candidates | 101 |
| 3.7.2 | A Messier Puzzle | 104 |
| 3.7.3 | The Proposed Diagnosis | 106 |
| 4. | Personal Identity | 111 |
| 4.1 | Introduction: The Facts of Life | 111 |
| 4.2 | Setting the Stage | 115 |
| 4.2.1 | A Context for Parfit's Argument | 115 |
| 4.2.2 | What Fission Might Show | 116 |
| 4.3 | The Argument and its Crucial Assumptions | 120 |
| 4.3.1 | Parfit's Fission Argument | 120 |
| 4.3.2 | Four Crucial Distinctions | 125 |
| 4.3.3 | Comments on these Distinctions | 127 |
| 4.3.4 | The Intrinsicness Premise | 129 |
| 4.3.5 | Summary | 129 |
| 4.4 | Two Unsuccessful Strategies | 130 |
| 4.4.1 | An Unsuccessful Attack on the Intrinsicness Premise | 130 |
| 4.4.2 | An Unsuccessful Defense of the Intrinsicness Premise | 131 |

| | | |
|-------|--|-----|
| 4.5 | Why is the Fission Argument so Compelling? | 134 |
| 4.5.1 | The Casewise Explanatory Difference Principle | 134 |
| 4.5.2 | The Casewise Explanatory Principle and the Method of Agreement | 136 |
| 4.5.3 | Fission and the Method of Agreement | 136 |
| 4.6 | How Absent Features can be Explanatory | 138 |
| 4.6.1 | Human Bodies and Borrowed Luster | 138 |
| 4.6.2 | Explaining Valuation | 139 |
| 4.6.3 | Exceptions, Norms and Local Adaptation | 142 |
| 4.6.4 | Prudential Concern in a World of Fission | 143 |
| 4.7 | Conclusion | 146 |
| 4.7.1 | Summary | 146 |
| 4.7.2 | Larger Lessons | 147 |
| 5 | Conclusion | 149 |
| 5.1 | Factive Thought Experiments: Galileo | 150 |
| 5.2 | Conceptual Thought Experiments: The Ship of Theseus | 154 |
| 5.3 | Valuational Thought Experiments: Personal Identity | 156 |
| 6 | Bibliographies | 161 |
| 6.1 | Bibliography of Works Cited | 161 |
| 6.2 | Bibliography on Galileo, Experiment and Thought Experiment | 189 |
| 6.3 | Bibliography on Personal Identity and Identity | 194 |
| 6.4 | Bibliography on Thought Experiment and Experiment | 229 |
| | Index | 251 |

Preface

0.1 REVISIONS MADE

The pages that follow represent a slightly revised version of my dissertation (completed in 1996 under the supervision of Robert Nozick, Derek Parfit and Hilary Putnam). Although I have not made major structural revisions, much of the text that appears below differs from what originally appeared in the dissertation. Some of the changes have been made in the course of preparing chapters for publication as articles; others have been made specifically with this manuscript in mind.

The changes made to chapter 1 are minor: examples have been slightly modified, terminology has been standardized, a few footnotes have been added or omitted, references have been updated, and material from (what appeared in the dissertation as) Appendix A has been incorporated into the text. (Appendices A, B, and C, which appeared in the dissertation, have been omitted from the published version of this work.) Elements of Chapter 1 have been incorporated into section IV of my “Exceptional Persons: On the Limits of Imaginary Cases.” *Journal of Consciousness Studies*, Vol. 5, Nos. 5-6 (1998): 592-610. The remaining revisions are specific to this volume.

The changes made to chapter 2 are rather more extensive: section 2.2.4 has been significantly rewritten, section 2.3.4 is entirely new, and a number of arguments have been refined throughout. The text that appears in this volume is almost identical to that published as “Galileo and the Indispensability of Scientific

Thought Experiment" in *British Journal for the Philosophy of Science*, Vol. 49, No. 3 (September 1998): 397-424.

Sections 3.1 through 3.6 of Chapter 3 have been edited slightly and section 3.7 has been significantly rewritten, all specifically for the purpose of this volume. Even with these changes, the chapter remains unsatisfactory: although the critical sections (3.5 and 3.6) strike me as quite sensible, the positive proposals (especially those in 3.4.2 and 3.7) neglect a number of important problems, and the text as a whole fails to take account of central metaphysical issues, particularly those raised in recent discussions of the problem of material constitution.

A number of arguments in Chapter 4 have been considerably revised in preparation for publication: Most of sections 4.3 and 4.5 have been extensively rewritten, and the rest of the chapter has been revised accordingly. Chapter 5 is virtually unchanged: A few references have been added, cross-references have been updated, terminology has been standardized, and stylistic infelicities have been corrected.

The volume also contains four bibliographies. The first, "Bibliography of Works Cited" (6.1) provides references for materials cited in the text itself. The three others, prepared with the research assistance of Dave Bzdak, offer comprehensive listings of post-1970 English-language books and articles (and, less comprehensively, earlier works) on the topics of: Galileo, Experiment and Thought Experiment (6.2); Personal Identity and Identity (6.3); and Thought Experiment and Experiment (6.4)¹.

0.2 SUMMARY OF CONTENTS

This study offers a novel analysis of the widely-used but ill-understood technique of thought experiment. Thought experiment, which is a process of reasoning carried out within the context of a well-articulated imaginary scenario in order to answer a specific question about a non-imaginary situation, is extensively employed in fields as diverse as physics, computer science, law, and philosophy. What I argue is that both the powers and limits of this methodology can be traced to the fact that when the contempla-

¹ I thank the Allan and Anita Sutton Fund of Syracuse University for providing me with a book subvention in support of this assistance.

tion of an imaginary scenario brings us to new knowledge, it does so by forcing us to make sense of exceptional cases.

The study has five chapters: an introduction, three case studies and a conclusion. My main contention is that certain patterns of features which coincide only fortuitously may nonetheless play a central role in the organization of our concepts, and that to the extent that imaginary scenarios involve disruptions of these patterns, our first-order judgments about them are often distorted or even inverted.

0.2.1 Chapter 1—Introduction

I begin the introductory chapter by identifying a certain class of exceptional cases with which the volume is primarily concerned. These are cases where some entity which falls within the purview of a theory lacks the characteristics (which I call the proxy characteristics) by which entities described by that theory are ordinarily identified². I contrast these sorts of theories (privileged-characteristic theories) with a second sort of theory (prototype- or paradigm-characteristic theories) where there is no such privileged characteristic³.

With this distinction in place, I suggest that there are two main sorts of strategies users of a theory can employ in accounting for exceptional cases. The first strategy is to use exceptional cases as a way of progressively narrowing the range of privileged characteristics, thereby ascertaining the theory's exception-driven norms; the second strategy is to use exceptional cases as evidence for the strength of the theory's core, thereby ascertaining the theory's norm-driven exceptions.

² So, for instance, according to some theory of sex, a person might fail to have the clothing, body type, social role, and emotional characteristics by which women are generally identified, but she might nonetheless be a (exceptional) *woman* because she has the requisite *privileged characteristics*, namely those which concern reproductive organs and chromosomal structure.

³ A theory of gender is a theory of this second sort. What it is to be an exemplary *female* is to exhibit some (maximal or near-maximal) subset of the characteristics which together characterize paradigm cases of femaleness—but no single characteristic is privileged such that possession of it alone is sufficient to make someone female. Correspondingly, an exceptional female is a female who exhibits few of these characteristics.

In later chapters, I argue that appeal to exceptional cases is indispensable if we wish to avoid mistaking accidental regularities for regularities which reflect a deeper truth about the world. But I show how the failure to attend to the distinction between theories with norm-driven exceptions and theories with exception-driven norms can lead to radically misguided conclusions.

0.2.2 Chapter 2: Galileo (First Case Study)

In chapter 2, I discuss the thought experiment with which Galileo is supposed to have refuted the Aristotelian view that heavier bodies fall faster than lighter ones⁴. Through a careful examination of this factive thought experiment⁵, I try to show that the guided contemplation of an imaginary scenario can provide us with new scientific knowledge in a way that argument alone cannot.

In the first part of the chapter, I explain the philosophical motivations for adopting the view that any scientific thought experiment can be reconstructed as a non-thought-experimental argument without loss of demonstrative force, and distinguish two versions of the position. The first—the dispensability thesis—concerns the replaceability of thought experiments; the second—the derivativity thesis—concerns their justificatory force. In the remainder of the chapter, I try to refute both of these theses. Through a detailed discussion of Galileo’s thought experiment, I show that the standard argumentative reconstruction of the case fails to capture its justificatory power, and I suggest reasons to think that any other argumentative reconstruction would fail in similar ways. I then argue that even if one were to provide an argumentative reconstruction that did almost perfectly capture the thought experiment’s demonstrative force, this would not show that the reason the thought experiment is successful is because—deep down—it is nothing more than an argument in disguise. I sug-

⁴ Imagine that a heavy and a light body are strapped together and dropped from a significant height. What would the Aristotelian expect to be the natural speed of their combination?

⁵ A *factive* thought experiment is one where the question at hand is: “What would happen?” Such thought experiments can be contrasted with what I call *conceptual* thought experiments, where the question is: “How, given an answer to the factive question, should we describe what would happen?” and with *valuational* thought experiments, where the question is: “How, given an answer to the conceptual question, should we evaluate what would happen?”

gest that, to the contrary, the success of the thought experiment is a result of the way in which it invites the reader’s constructive participation, describes particulars in ways that make manifest practical knowledge, and describes an imaginary scenario wherein relevant features can be separated from those that are inessential to the question at issue.

0.2.3 Chapter 3: The Ship of Theseus (Second Case Study)

In Chapter 3, I discuss the story of the Ship of Theseus⁶. I try to show that standard interpretations of the puzzle have drawn the wrong sorts of general conclusions from the particular case by assuming that the proper way to understand artifact identity is to allow the exceptions to drive the norms, rather than allowing the norms to drive the exceptions. I argue that it is only against a background norm of intrinsically-determined identity that we can make sense of local instances of extrinsically-determined identity⁷.

The story of the Ship of Theseus presents us with a case where a process which is ordinarily identity-preserving is instead entity-creating. I discuss a number of attempts to dissolve (van Inwagen and Parfit) and solve (Hirsch and Nozick) the puzzle, and I contend that these proposed resolutions misidentify the case’s implications by locating the extrinsic determination in the identity criteria themselves. What I argue instead is that the best way to account for such cases is to locate the extrinsic determination in the processes by which identity is generally preserved, for to do so is

⁶ Suppose there is a ship whose planks are gradually replaced over time, while the discarded planks are collected in a barn. Suppose that the discarded planks are then reassembled in the shape of the original ship. Which one is the original: the ship which has undergone constant repair, or the ship made from the disassembled and reassembled planks?

⁷ That is, cases where whether x is identical with y depends on something other than the spatio-temporal and causal relations between x and y .

to help ourselves to a constraint which is already at play when we think about identity over time.

0.2.4 Chapter 4—Personal Identity (Third Case Study)

In Chapter 4, I discuss the case of fission⁸, which has been taken to establish that personal identity is not what matters for rational prudential concern. While accepting that in some counterfactual⁹ cases rational prudential concern can hold in the absence of identity, I argue that what explains the rationality of prudential concern is always identity. I do so by showing that the feature that explains or justifies or makes rational a relation can be a different feature from the one that underpins it as a necessary condition. And I show this by arguing in detail that cases such as fission need to be understood as norm-driven exceptions.

I suggest that the reason the fission argument seems so compelling is because of its tacit reliance on an ostensibly undeniable principle concerning the assignment of explanatory force. I contend that the principle does not apply to the fission case in the manner that the argument requires, because the argument neglects the ways in which certain patterns of features that coincide only fortuitously may nonetheless play a central role in the organization of our concepts. To the extent that imaginary scenarios involve disruptions of these patterns, our first-order judgments about them are often distorted or even inverted.

0.2.5 Chapter 5—Conclusion

In the final chapter, I attempt to tie together the three case studies by showing that exceptionality plays a role in their evaluation at

⁸ Suppose it were possible to transplant your brain into another body nearly identical to your own, and suppose that either the left or the right hemisphere of the brain would be sufficient to support full preservation of all mental properties. Suppose that your right and left hemispheres are successfully transplanted into two different bodies, such that each has all of your memories, beliefs, desires, etc. You are identical to neither continuer. What sort of (rational) relation should you bear towards them? What does this show?

⁹ In the introduction, I distinguish between imaginary cases which are *hypothetical* (for all we know, such a situation might indeed obtain) and cases which are *counterfactual* (without looking at the world, we know the situation is imaginary).

three distinct levels: (a) in the patterns of classification of certain states of affairs as normal or exceptional; (b) in the ways in which the particular exceptional (that is, unusual) situation described by the scenario should be accounted for; and (c) in the ways in which this exceptionality is accounted for at the level of explanation.

In the Galileo example: (a) the Aristotelian comes to see that the cases he has been taking to be exceptional cases (cases where bodies fall with the same speed) are in fact typical, and vice versa; (b) the case asks him to contemplate an object which is exceptional in the sense that its criteria of entification are far from obvious, but which the norms of science require him to account for by means of exception-driven norms; (c) because exceptions in explanation are norm-driven, the case forces him to look at what once seemed to be normal through the lens of the new normal cases.

In the case of the Ship of Theseus, (b) the unusual case presents us with an instance where a process which is ordinarily entity-preserving is instead entity-creating. But here (a) it is not true that the proper conclusion to draw is that such processes are in general entity-creating (or that they provide dubious mechanisms for identity-preservation). For (c) if cases like Theseus were the norm rather than the exception, it wouldn't make sense even to speak of identity-candidacy for ships. Explanation goes via normal cases; they provide the background against which we make judgments about the atypical.

Finally, (b) the case of fission is exceptional in the flat-footed second sense; fission is meant to be extraordinary, not commonplace. And, (a) as I argue extensively in the body of chapter 4, the appropriate way to account for this case is to recognize that for it to make sense to us, it must be exceptional; our ability to judge value in such cases is parasitic on our ability to judge value in ordinary cases. What this means is that (c) when we contemplate such counterfactual cases, our very judgments are filtered through the background patterns of fortuitous coincidence which play central roles in the organization of our conceptual scheme; there are certain things we simply cannot learn through the contemplation of imaginary cases.

This brings us full circle to the Galileo case. There, we saw that the thought experiment is able to bring new knowledge to the Aristotelian by allowing him to see all cases involving falling bodies through the lens of cases where bodies fall with the same natu-

ral speed. This is what explains the thought experiment's power in the Galileo case. And in the fission case, precisely the same feature explains the thought experiment's limitations. We cannot suddenly step outside the framework within which we ordinarily make sense of what we value and believe. The way the world is and the way we see the world are too deeply intertwined for this to be possible.

0.3 ACKNOWLEDGMENTS

The acknowledgments section of my dissertation occupied some 5 pages, and was filled with anecdote and detailed expressions of gratitude. Though I stand by the sentiments expressed therein, it seems out of place to repeat them in this public setting. I trust that those listed below will understand that my appreciation has not waned.

Versions of (what have ended up as parts of) Chapters 1, 2, 3 and 4 were presented as papers before audiences at the Second Annual Harvard-MIT graduate student conference (Chapter 1); at the MEPHISTOS Conference in the Methodology, Philosophy, History and Sociology of Science, the Harvard University Nth Year Seminar, and the Department of Philosophy at Cornell University (Chapter 2); at the Department of Philosophy at Reed College (Chapter 3); and at the Departments of Philosophy at Harvard, Cornell and Syracuse Universities (Chapter 4). For comments on earlier versions of these chapters, I am grateful to the members of those audiences, and to Andrew Botterell and Steven Gross (Chapter 1); Richard Boyd, James Robert Brown, Michael Glanzberg, Steven Gross, Ned Hall, Norman Kretzmann, Thomas Kuhn, John Murdoch, John Norton, W.V. Quine, Sherri Roush, Simon Saunders, Roger Shepard, Alison Simmons and Zoltán Gendler Szabó (Chapter 2); John Hawthorne, Teresa Robertson, Ted Sider, and Zoltán Gendler Szabó (Chapter 3); and Richard Arneson, Richard Boyd, Michael Della Rocca, Jennifer Dworkin, Warren Goldfarb, John Hawthorne, Harold Hodes, Terence Irwin, Mohammed Ali Khalidi, Norman Kretzmann, Scott MacDonald, Elijah Millgram, Richard Miller, Ian Proops, Carol Rovane, Sydney Shoemaker, Ted Sider, Susanna Siegel, Jason Stanley, Eleanor Stump, Nicholas Sturgeon, Zoltán Gendler Szabó and Jennifer Whiting (Chapter 4).

During my graduate career, I was fortunate to have received funding from the Mellon Foundation (1989-90, 1990-91 and 1994-95) and the National Science Foundation (1991-92, 1992-93 and 1993-94). I am grateful to both of these organizations for their financial support, and I hope that both will continue to have the will and the resources to offer similar support to others in the future. Since completing the dissertation, I have had the pleasure of being employed by the Department of Philosophy at Yale University (1996-1997) and as the Allan and Anita Sutton Distinguished Faculty Fellow in Philosophy at Syracuse University (since 1997); both settings have provided outstanding working environments, from which I have benefited greatly.

Seven people deserve special mention. The greatest influence on this work has come from the three members of my dissertation committee, Robert Nozick, Derek Parfit and Hilary Putnam, each of whom served as an exemplary teacher, mentor, and advisor throughout my graduate school career, and has continued to do so ever since. More recently, in revising the dissertation for publication, I have benefited immeasurably from conversations with my extraordinary colleagues Ted Sider and (most especially) John Hawthorne, and with my husband Zoltán Gendler Szabó. Finally, I am grateful to David Bzdak for his assistance in preparing the exceptionally comprehensive bibliographies, and for his outstanding work in producing the volume's index.

CHAPTER 1

Introduction

Thinking about imaginary cases can help us learn new things about the world. This simple fact is both a commonplace, and a puzzle. It is a commonplace because it is undeniable that imaginary test cases play a central role in our investigation of the world—in legal reasoning, in linguistic theorizing, in philosophical inquiry, in scientific exploration, and in ordinary conversation. And it is a puzzle because it is *prima facie* surprising that thinking about what there isn’t and how things aren’t should help us to learn about what there is and how things are.

The goal of this study is to show that this simple fact is both less of a commonplace and less of a puzzle than it might initially seem. It is less of a commonplace because the failures of this technique are far more prevalent than its enthusiasts would admit. And it is less of a puzzle because what accounts for its success is far less mysterious than its detractors would concede. Moreover, the explanation for both its success and its failure is the same. Both the powers and the limits of the technique can be traced to the fact that when the contemplation of an imaginary scenario brings us to new knowledge, it does so by forcing us to make sense of an exceptional case.

1.1 EXCEPTIONAL CASES

1.1.1 Characterization of “Exceptional Case”

1.1.1.1 Exceptional Cases in Theories with Privileged Characteristics

What I mean by “exceptional case” can be characterized more precisely as follows. (I will give an example of what I mean in the next paragraph.) Suppose that there is some theory which, in general (i) accounts for entities or situations that are identified—by users of the theory—under more than one description or on the basis of more than one characteristic. Now take two of these descriptions such that: (ii) It is not conceptually necessary that these descriptions co-vary, and (iii) one description has privileged status within the theory, such that what the theory purportedly aims to do is to say something about those entities or situations under that description. Let us call this the “privileged description” and call the corresponding characteristic the “privileged characteristic.” Suppose that (iv) in a majority of cases, the entities or situations in question can also be picked out by some proxy characteristic which is generally a concomitant of the privileged characteristic. Then (v) An exceptional case is then one where there is some entity or situation which falls under the purview of the theory, but which lacks the proxy characteristic¹.

As an example, take the notion of velocity described in Jean Piaget's famous series of experiments concerning children's conceptions of motion and speed (see Piaget 1964/1970 and discussion in Kuhn 1964). In the experiments, children were instructed to observe a pair of toy cars that were moved uniformly along a straight line towards some pre-specified goal. One of the cars was blue, the other red, and at the completion of each trial, the children were asked which of the cars had moved faster.

¹ I do not mean to suggest that this is the only way in which a case might be exceptional. Many of the scientific examples I go on to discuss in the text below are exceptional in a much more mundane sense; they offer instances of phenomena that are atypical because they manifest an unusual* configuration of properties. The characterization of “exceptional” that I have just offered is a special case of this more general phenomenon. A discussion of the ways in which exceptional cases of this more special sort are connected to exceptional cases in general can be found in the concluding chapter.

*Obviously, what is usual or unusual is so with respect to some set of background assumptions; nothing is unusual *simpliciter*.

The children confronted three sorts of trials. In the first sort, both cars moved the same distance but one took longer to do so. So, for instance, the blue car and the red car might leave the starting line together, but the red car would move at a steady pace of p along the entire course, while the blue car moved at a steady pace of $p+1$. In the second sort of trial, both cars moved for the same amount of time but one covered a greater distance. So, for instance, the blue car and the red car might leave the starting line together, the blue car moving at a pace of $p+1$ and the red car at a pace of p , and both would stop after some specified period of time. In the third sort of trial, neither the times nor the distances were the same. So, for instance, the blue and red cars might set off together, with speeds as before, but the blue car might stop after a certain amount of time, while the red car continued on a bit longer. After each trial, Piaget asked the child which car had gone faster, and why².

The children, aged about five, tended to describe as faster the car which reached its goal first; this was so even when they recognized that the “slower” car had traveled a greater distance during the same period of time. Thomas Kuhn suggests that the child's response here manifests what he calls the goal-reaching criterion for the application of faster (Kuhn 1977, 244). That is, the child uses the term “faster” to describe the object that reaches its goal first, regardless of differences in the length of path covered, or the relative rates of motions of the vehicles.³ But in addition to the

² The experiments I discuss are described in Kuhn 1977, 243-256. For complete text, see Piaget 1946/1970; the examples in question can be found in chapters 7 and 9.

³ So, for instance, in the following example, the red car and the blue car both travel for the same amount of time, but the red car starts from a point well behind the blue car, and catches up with it.

P(piaget): Same speed or not?

C(hild): Both the same.

P: Do they do the same distance?

C: No, one starts here and the other there.

P: The same length?

C: No, that blue one is shorter.

P: Were they both going just as hard?

C: Yes.

P: One of them wasn't faster than the other?

C: No.

P: Which one went a long way?

C: Red.

P: And what if I said to you that one was going harder than the other?

C: Well, I think they went at the same speed. (Piaget 1946/1970, 158).

"goal-reaching" criterion, children of this age also apply a second criterion to determine which of two objects is moving more quickly. Kuhn refers to this criterion as "perceptual blurriness." One object is "perceptually blurrier" than another if its more rapid motion can be directly observed⁴.

These two criteria—goal-reaching and perceptual blurriness—coexist unproblematically in the majority of cases. In most of the trials presented by Piaget, and in most daily experience, faster-moving objects reach their goals earlier than slower-moving objects⁵. However, as Kuhn notes:

Not very often (or the children could not have preserved the concept for so long) but occasionally nature will present a situation in which a body whose directly preserved speed is lower nonetheless reaches the goal first. In this case the two clues conflict; the child may be led to say that both bodies are "faster" or both "slower" or that the same body is both "faster" and "slower." That experience of paradox is the one generated by Piaget in the laboratory (Kuhn 1977, 244-245).

That is, in most cases, the object that reaches its goal first is the object that moves most quickly along the designated path. Occasionally, however, the faster-moving object will begin its journey much later than the slower-moving object, or will trace a much longer path. In this circumstance, the object in question presents an exceptional case, which can be characterized as follows: (i) Users of the theory identify the faster participant (that is, the participant that has the greatest average speed) by examining how the participants look when they are in motion (perceptual blurriness), by checking to see which participant reaches the finish line first (goal-reaching), and by calculating the participant's mean speed over the course. (ii) It is neither conceptually nor nomologically necessary that these criteria coincide: the "blurriest" participant

⁴ So, for instance, when the case just described was modified so that the red car began moving much later than the blue car, and correspondingly moved much more rapidly, the same child responded as follows:

P: Did one go more quickly than the other?

C: The red.

P: How did you find that out?

C: I watched it!

⁵ This suggests that "The Tortoise and the Hare" teaches cognitive as well as moral lessons.

will not always reach the finish line first, nor will the blurriest participant or the participant who reaches the finish line first always be the participant who completes a given course in the shortest amount of time. (iii) What the theory really concerns is this last characteristic, namely, average velocity; this is the "privileged characteristic." (iv) In most cases, the participant that has the greatest average velocity will also be the participant that is "blurriest," as well as being the participant that reaches the finish line first. Hence either of these characteristics can serve a proxy for the privileged characteristic. (v) An exceptional case is one where the participant that is "blurriest," or the participant that reaches the goal first is not in fact the participant with the greatest average velocity.

We might also consider, more briefly, a second example. A theory of sex (i) accounts for persons who are identified—by users of the theory—on the basis of more than one characteristic; for instance, women are identified on the basis of reproductive capacities, chromosomal structure, manner of dress, body shape, social role, and emotional tendencies. (ii) It is not conceptually or nomologically necessary that these descriptions co-vary: a person may have a uterus and wear men's bathing trunks, or she may work as a truck driver and have a close and caring emotional relationship with her children. (iii) One description has privileged status within the theory, such that what the theory purportedly aims to do is to say something about persons under that description. In this case, the "privileged description" is the biological description, and the corresponding "privileged characteristics" are, let us say, those that concern chromosomal structure. (iv) In a majority of cases, persons can be picked out by some proxy characteristic that is generally a concomitant of the privileged characteristic. So, for instance, a person wearing a skirt is likely to have the chromosomal structure in question, as is a person who is a nursery school teacher. (v) An exceptional case is then one where some person who falls under the purview of the theory as a result of having the requisite chromosomal structure lacks the proxy characteristics: for instance, someone who is unable to bear children, wears Harley-Davidson T-shirts, is highly muscular, is romantically involved with a number of women, works on an oil rig, and has a violent temper.

1.1.1.2 Exceptional Cases in Theories without Privileged Characteristics

Both of the examples described so far involve theories where one characteristic has privileged status (as required by (iii)). But the term “exceptional case” may also apply to cases within a theory that lacks characteristic (iii), that is, a theory that concerns entities that have some (proper or improper) subset of the cluster of characteristics that together typify the entities that the theory concerns. In the case of such a theory, we might wish to modify (iii) as follows: (iii') some cluster of properties characterizes entities described by the theory, such that what the theory purportedly aims to do is to say something about those entities or situations that exhibit some significant portion of these characteristics. We will call entities that exhibit the designated properties “paradigm cases”.⁶⁷ In keeping with this adaptation of (iii), we might corre-

⁶ Paradigm cases may exhibit all of the designated properties, or as many as normally possible if it is not normally the case (or not possible) that all of the characteristics in question are held conjointly by a single entity or occur conjointly in a single situation. My characterization here is intentionally sketchy, as I am hoping to remain neutral among the following sorts of cases. (The list that follows is not intended to be exhaustive.) Suppose Xs are characterized by having some (proper or improper) subset of the properties a, b, c, and d. It may be that:

- Paradigm X has all four of these properties.
- Paradigm X has some particular trio of these properties; no X does (or could) have all four.
- Paradigm X has some particular trio of these properties; some Xs do (or could) have all four.

Note that speaking of a “paradigm” may be misleading in cases of the following sort:

Xs have properties a and b

Ys have properties b and c

Zs have properties a and c

Xs, Ys, and Zs are the only instances of the entities in question—call them Ns. Here we have two choices:

(1) we might say that there is a paradigm case of being an N, namely being an entity that has a, b and c; or

(2) we might say that there is no single paradigmatic N, but rather, that there are three prototypical Ns, namely: Xs, Ys, and Zs.

There are two problems with the first reply: (a) In some sense, the entity with a, b and c—even if it were to exist—would be extremely atypical, in that (i) it resembles no actual entity, and (ii) it differs from all actual entities in a straightforwardly

Continued on next page

spondingly replace (iv) by (iv')—in a majority of cases, the entities or situations in question can be picked out by many of the characteristics which together characterize paradigm cases—and (v) by (v')—an exceptional case is one where there is some entity or situation that falls under the purview of the theory, but that cannot be picked out in the way (iv') requires⁶⁸.

A theory of gender is a theory of this sort. (iii') Some cluster of properties characterizes entities described by the theory, such what the theory purportedly aims to do is to say something about those entities or situations that exhibit some significant portion of these characteristics. So, for instance, females are persons who occupy some particular set of social roles, exhibiting a range of characteristics that epitomize being female. Persons who exhibit some (maximal or near-maximal) subset of the designated properties are paradigm instances of females⁶⁹. (iv') in a majority of cases, females

specifiable way, namely, it has three of the properties that appear only pair-wise in all actual instances (and the latter might itself be what we seek). (b) There are certain problems that arise with imaginary cases that do not arise with actual cases, such that taking an imaginary case as a paradigm may lead to certain sorts of problems. For these reasons, it makes sense to go with the second reply.

⁷ Note that I am using the terms here somewhat differently from how they are used in the psychological literature on paradigm or prototype views of concepts. For an overview these debates, see Komatsu 1992, Medin and Goldstone 1990, Medin and Smith 1984, or Smith and Medin 1981. For a classic presentation of the prototype view, see Rosch and Mervis 1975 (following Wittgenstein 1953); for recent criticism, cf. Fodor 1998. For related further discussion, see the papers collected in Margolis and Laurence 1999, Neisser 1987, or Rosch and Lloyd 1978.

⁸ Either because the entity has too few of the relevant features, or because it has too many.

⁹ In cases of the sort described in footnote 6, we would then say: (iv") in a majority of cases, the entities or situations in question can be picked by some sufficiently large subset of the characteristics that together make up the disjoint set of characteristics that can be found among prototypical cases. (v") An exceptional case is one where there is some entity or situation that falls under the purview of the theory, but that cannot be picked out in the way (iv") requires.

¹⁰ It is unclear how comprehensive this subset could be; one frequently-made critique of traditional gender roles has been that it is impossible for a woman to satisfy simultaneously all of the requirements of being a “successful” female; she cannot be simultaneously delicate and capable of doing all the housework, morally upstanding and sexually available, deferential in decision-making and capable of taking charge, etc.

can picked out by many of the characteristics that together characterize paradigm cases. (v') An exceptional case is one where some person falls under the purview of the theory but cannot be picked out in the way (iv') requires, that is, some case where there is a person who is—according to the theory in question—a female, but who bears few of the attributes associated with paradigm instances of females¹¹.

1.1.2 Ways of Accounting for Exceptional Cases

There are two main ways a theory can account for exceptional cases; or, to speak a bit more precisely, there are two main sorts of strategies users of a theory can employ in accounting for exceptional cases. The first strategy is to use exceptional cases as a way of progressively narrowing the range of privileged characteristics. So, for instance, suppose entities accounted for by the theory in question generally have characteristics *a*, *b*, *c*, *d*, and *e*. Suppose further that as an exceptional case, some entity is found that falls within the purview of the theory, but which has only *b* and *d*. It then follows that no characteristics other than *b* and *d* can be privileged characteristics in the sense that they necessarily belong to any entity that falls within the purview of the theory.

Such an attitude towards exceptional cases involves using them as test cases to ascertain necessary and sufficient conditions. The conclusion then drawn is that even in non-exceptional cases, the characteristics that really matter are those that are present in the exceptional cases as well. According to this strategy, one uses exceptional cases to ascertain the theory's exception-driven norms. The exceptions drive interpretation of the norms; what is taken to matter about normal cases is whatever it is that they have in common with exceptional cases. So if normal cases of *X*s have *a*, *b*, *c*, *d*, and *e*, but exceptional cases have only *b* and *d*, then on this strategy, we would say that all that matters even in normal cases of *X*-ness is that *b* and *d* are present.

The second strategy is to view exceptional cases as evidence for the strength of the theory's core. So, for instance, suppose again

¹¹ It might also be that an individual who bears all the attributes would be exceptional, but the exceptionality would be of a different sort. (See also footnotes 1, 6 and 8.)

that entities accounted for by the theory in question generally have characteristics *a*, *b*, *c*, *d*, and *e*, and suppose further that some entity is found that has only *b* and *d*, but which nonetheless seems to fall within the purview of the theory. According to the second strategy, the proper thing to say about the entity in question is that it falls within the purview of the theory, but only because it is similar in certain crucial ways to more typical instances of entities that the theory describes. Under this strategy, one uses exceptional cases to ascertain the theory's norm-driven exceptions. The norms drive interpretation of the exceptions. On the basis of this strategy, one concludes that what it is that allows the exceptional cases to be cases at all is that they have enough in common with the normal cases. The reason the entity with only *b* and *d* counts as an *X* is because in general, *b* and *d* are found only in association with *a*, *c*, and *e* as well.

So, for instance, suppose again that both women and females generally have characteristics *a*, *b*, *c*, *d*, and *e*, where these are: *a*: wears thus-and-such sorts of clothing; *b*: has thus-and-such chromosomal structure; *c*: occupies thus-and-such social role; *d*: has thus-and-such body parts; *e*: tends to respond to stressful situations in thus-and-such a way. And suppose that some person is found who is held to be both female and a woman, but who has only *b* and *d*. That is, the person in question has the designated chromosomal structure and body parts, but none of the social characteristics ordinarily associated with being a woman or being female. According to the first strategy, such a case would show that all that matters about being a woman or being female is having the requisite genetic and bodily characteristics. According to the second strategy, such a case would show that the reason such genetic and bodily structures “count” in making one female or a woman is because, in general, they are associated with the social characteristics enumerated above.

1.1.3 Patterns of Accounting

It seems to me (as I hope it seems to my reader) that the first strategy (exception-driven norms) is appropriate when we are talking about what I am calling women¹², whereas the second strategy

¹² For the sake of this discussion, I am not considering more radical critiques which suggest that the categories I am calling man and woman—that is, the biological categories by which classification is made throughout the animal kingdom—are themselves socially constructed.

(norm-driven exceptions) is appropriate when we are talking about what I am calling females. That is, the proper conclusion to draw from the fact that someone can be a woman if she has only *b* and *d* is this: even in instances where the person in question also has *a* and *c* and *e*, the factors that really matter in making her a woman are *b* and *d*. By contrast, the proper conclusion to draw from the fact that someone can be a female if she has only *b* and *d* is not that *b* and *d* are all that matter even in ordinary cases. Rather, the proper conclusion to draw is that it is only because *b* and *d* are generally found in conjunction with *a*, *c* and *e* that we can—by courtesy as it were—include the *b-d* individual within the purview of the theory.

What I want to ask in this section is: is there anything systematic that can be said about what makes the first strategy appropriate in the first case, and the second strategy appropriate in the second case? My suggestion is that there are two tests that can be applied.

The first test employs the distinction made previously, between theories that attempt to enumerate necessary and sufficient conditions for being an *X*, and theories that can be properly said to characterize *Xs* by identifying characteristic instances of *Xs* and enumerating their features. In theories of the first sort, exceptional cases are appropriately accounted for by means of the first strategy; these are theories with exception-driven norms. In cases of the second sort, they are appropriately accounted for by means of the second strategy; these are theories with norm-driven exceptions.

On reflection, this should not be surprising. A theory that characterizes *Xs* on the basis of necessary and sufficient conditions must account for every case that falls under its purview on the basis of these conditions. So if exceptional-*X* manifests only some of the characteristics associated with typical-*X*, it follows that even for typical-*X*, the only *X*-making characteristics can be those that are present in exceptional-*X* as well. By contrast, a theory that characterizes *Ys* on the basis of paradigm instances must have instances that are paradigmatic for the theory to be a theory of anything at all¹³. So if exceptional-*Y* manifests only some of the characteristics associated with typical-*Y*, it does not follow that the only *Y*-making characteristics are those that are present in both typical-*Y* and exceptional-*Y*. On the contrary, what follows is that

¹³ For the sake of simplicity, I am temporarily ignoring family-resemblance cases of the sort described above. (See footnote 6.)

even for exceptional-*Y*, the *Y*-making characteristics are those that are present in typical-*Y*.

The reason for this is brought out by the second test. The second test involves asking the following question: what would happen if all cases were like the exceptional case; would there still be *Xs*?¹⁴ In theories with exception-driven norms, the answer is “yes;” in theories with norm-driven exceptions, the answer is “no.” So, for instance, to return to the example above: both women and females generally have characteristics *a*, *b*, *c*, *d*, and *e*, and the exception we are considering is one where the person in question has only *b* and *d* (the biological characteristics), and lacks *a*, *c* and *e* (the social characteristics)¹⁵. We now ask: what if every case were like this one? And the answer is: there would still be women; but there would not be females¹⁶.

In short, what we do when we characterize what it is to be a woman in the sense I am describing is to enumerate a set of necessary and sufficient characteristics. So even all cases were to manifest only the bare minimum of such characteristics, there would still be women. By contrast, what we do when we characterize what it is to be a female in the sense I am describing is to identify a set of attributes that are ordinarily correlated with one another. So if all cases were to manifest only the bare minimum of such characteristics, there would be no females. Or, more precisely, there would be no category “female” to which the persons who are currently classified as females would belong. With the breaking

¹⁴ As Kant notes, certain practices are, if generalized, self-undermining. See, for instance, his discussion of lying in the *Groundwork* at Kant 1785/1964, 29/71 and 55/90.

¹⁵ As before, these can be understood as follows: *a*: wears thus-and-such sorts of clothing; *b*: has thus-and-such chromosomal structure; *c*: occupies thus-and-such social role; *d*: has thus-and-such body parts; *e*: tends to respond to stressful situations in thus-and-such a way.

¹⁶ I may seem to be stacking the deck unfairly, in that I am taking woman to describe a biological category and female to describe a social category. One might say: surely if all there were were instances of persons with *a* and *c*, there would not be women. My reply: Indeed—but if there were even one case of a person who had only *a* and *c* without *b* and *d*, that person would not be a woman. The question is: what would happen if all cases were like this exceptional case, where the exceptional case in question describes an entity that falls under the purview of the theory?

down of a pattern of correlated characteristics would come a breaking down of the concept that describes such a pattern.

1.1.4 Application to the Question at Hand

As the reader may recall, the reason I have been discussing exceptional cases is because they play a role in the central contention of the study, which is that both the powers and the limits of the contemplation of imaginary scenarios can be traced to the fact that when such contemplation brings us to new knowledge, it does so by forcing us to make sense of an exceptional case. And I contended further that this diagnosis explains both the powers and the limits of this technique. The question that naturally arises is: why is this so? My answer is three-fold. First, thinking about exceptional cases can lead us to a reconfiguration of our conceptual commitments, allowing us to organize information in a way that renders it newly meaningful. Moreover, exceptional cases are good test cases; they help prevent us from mistaking accidental regularities for regularities that reflect a deeper truth about the world. But, third, exceptional cases are dangerous; if we fail to keep straight the distinction between theories with norm-driven exceptions and theories with exception-driven norms, we are likely to draw radically misguided conclusions.

The last of these claims, I think, explains the adage that hard cases make bad law. The reason hard cases make bad law (that is, bad normative law) is because laws aim to provide normative guidance in ordinary cases. But since hard cases are exceptional cases, taking them as exemplary will lead to a misfit between theory and everyday practice. If we are trying to determine the age at which children should enter kindergarten, we should not make our decision by looking at John Stuart Mill; if we are trying to decide penalties for theft, we should not focus on a case where a man steals medicine to save his dying wife. We can make sense of exceptional cases here, but only against a background of normalcy: John Stuart Mill may go to kindergarten, but if he does, he will be precocious; the medicine-stealer may be a thief, but if he is, he is a thief who faces extenuating circumstances.

It is the second of these claims that explains the (not-quite) adage that hard cases make good science¹⁷. The reason hard cases

make good (descriptive) science is that science aims to provide systematic theories that explain not only all actual, but also all physically possible phenomena within their purview¹⁸. If a mere inventory of what there is were the sole goal of scientific theories, naïve regularity theories would capture what it is that we mean by physical law. But they do not¹⁹. Even if there is not enough radium on

(necessarily) things like classificatory biology or cultural anthropology. Of course, even here, hard cases might be a terrible place to start if one is trying to develop a theory. (This is a point about the epistemology of discovery and the capacities of human beings for pattern recognition.) But I am concerned with the status of hard cases as test cases. (This is a point about the epistemology of justification and the standards which we assume scientific theories will meet.)

¹⁸ Cf. Aristotle: "In all disciplines in which there is systematic knowledge of things with principles, causes, or elements, it arises from a grasp of those: we think we have knowledge of a thing when we have found its primary causes and principles, and followed it back to its elements" (Aristotle, *Physics* 184a11-15 in Ackrill 1987). Of course, a scientific theory might respond to apparent counterexamples by refining the range of cases which it purports to cover.

¹⁹ A naïve regularity theory is a theory which picks out as laws of nature those statements which are (a) universally quantified; (b) true; (c) contingent; and (d) contain only logical connectives, logical quantifiers, and non-local empirical predicates. (See Armstrong 1983, 12; the characterization in question comes from Molnar 1969.) Following Armstrong, we might call the uniformities picked out by such a characterization Humean uniformities or Humean regularities.

In identifying Humean uniformities with laws of nature, naïve regularity theories face a cluster of problems. (The following discussion is simply a summary of Armstrong 1983, 12-13.) First, they face extensional difficulties in that they pick out uniformities which do not seem to be laws of nature, so they take as sufficient for being a law of nature a characteristic (being a Humean uniformity) which is in fact not sufficient. Second, they face oversimplification difficulties in that they identify the content of laws with the content of uniformities, whereas there are cases where it seems that a gap exists between a law and the manifestation of that law, that is, between the content of a law and the content of the uniformity which is its manifestation. Finally, they face intensional difficulties, in that even if it is assumed that the content of the law and the content of the uniformity are identical, it might still be that the law has properties which the manifestation lacks. (For detailed arguments establishing the validity of these criticisms, see Armstrong 1983, 1-73. Regardless of what one thinks of the positive characterization he offers of his own theory in the second part of the book, his negative arguments against naïve regularity theories are, I think, quite convincing.)

¹⁷ By "science" here, I mean things like plain old-fashioned descriptive physics, not
Continued on next page

earth to fill a box as large as the Pentagon, we still want our theory to tell us how much radioactivity it would produce. And if a new theory of planetary motion fails to predict the motion of a new heavenly body that appears on the scene, so much the worse for the theory.

Whether hard cases make good or bad philosophy is a much more difficult question. Like normative legal theories, philosophical theories are often concerned with providing normative guidance in ordinary cases. But like scientific theories (indeed, even more than scientific theories), philosophical theories aim to provide us with accounts not only of how things are, but also of how things can and cannot be²⁰. This might seem to suggest a simple division: philosophical theories that are normative are likely to have difficulties with exceptional cases, whereas those that are descriptive are likely not to. Unfortunately, the distinction is not so simple.

1.2 IMAGINARY CASES

Because the world is neither maximally replete nor effortlessly navigable, all legal and scientific and philosophical investigation inevitably involves the contemplation of cases that are imaginary. Law school textbooks are brimming with hypothetical cases which are the central diet of legal education. Elementary physics texts are replete with postulated examples that ask the reader to imagine performing certain experiments, and to predict their outcomes. And elementary philosophy texts overflow with passages like this one, taken from the opening pages of Jonathan Dancy's introductory text on epistemology:

[In thinking about what we mean by “knowledge”] we do not need to rely on actual mistakes in the past. For our purposes, possible mistakes will do just as well. This can be seen in the ethical case. An imaginary example can be so described that I am willing to say that the action it recounts is good. And that judgment of mine is as much universalizable, as binding on my future judgments about relevantly similar cases, as if the

²⁰ “Experience tells us...what is, but not that it must necessarily be so, and not otherwise. It therefore gives us no true universality; and reason, which is so insistent upon this kind of knowledge, is therefore more stimulated by it than satisfied” (Kant 1787/1929, B2/A2).

example had been real rather than imaginary. Similarly, an imaginary case in which I would claim to know that p, but where p is false, will succeed in preventing me from claiming to know that p in a new case which is not relevantly (i.e., discernibly to me) different. So imaginary cases are as effective in the argument as actual ones. (Dancy 1985a, 13-14).

That is, Dancy suggests that when an imaginary case is indiscernibly different from an actual case, our responses to that case are as relevant as our responses to an actual case would be²¹.

I suggest the following spelling-out of Dancy's criterion: if, for all we know, the case presented to us might be actual, then our responses to that case are to be treated as we would treat responses to an actual case²². So, for instance, there is little if any difference between deciding whether an actual child raised as a Jehovah's Witness should be excused from the classroom during the Pledge of Allegiance, and deciding whether, if there were a child raised as a Jehovah's Witness in the class, that child should be excused from the classroom during the Pledge of Allegiance. This is true of any realistic normative case, where the situation described may, for all we know, be actual, and the judgment in question is about what should be done or said. In such cases, whatever it is that explains the informativeness of thinking about actual situations in order to decide the proper application of principles or concepts explains the informativeness of thinking about imaginary situations.

Of course, thinking about a situation—actual or hypothetical—is not the same as confronting a situation, and it may be that there are issues raised by reasoning as such as a way of determining answers to normative questions. Moving the case from the world to the mind “cleans it up” in certain ways, and allows it to be iso-

²¹ It is interesting to note that Dancy is one of the main contemporary opponents of the technique of thought experiment in ethics. (See Dancy 1985b, as well as his discussion of “switching arguments” in Dancy 1993, 64-66.)

²² But cf. Nozick: “[I]f principles are only supposed to cover the cases that will, would, and could arise, then before the fact, if it is thought such a case is impossible (that the situation, motivations, or whatever that would lead to it could not arise or succeed), it might not be considered a relevant counterexample to that or any principle. But once it is discovered that [such and such] can [occur]—because it did—then the principle P that [failed to account for it] is refuted” (Nozick 1993, 38).

lated from the network of events of which it is a part. But this is equally true of cases that are drawn from experience and of cases that are merely described. Still, it might seem—even for realistically described normative cases—that the difference between an imaginary and an actual situation about which we are called to make a judgment is that the latter carries with it a certain sort of urgency: some state of affairs that is actually out there, and we need to make a decision about how it is properly to be evaluated. An imaginary situation, by contrast, is purely academic: failure to make a judgment about it, or to make a proper judgment about it, might leave us with less knowledge in some scientific sense, but it would not—at least not directly—leave us less able to negotiate the world as it actually is.

Although this is a difference, I do not think it is a deep one—certainly not deep enough to drive a wedge between the class of cases that are actual and the class of cases that, though for all we know might be actual, as a matter of fact are not. In the first place, judgments about imaginary cases may help us in making judgments about actual cases, and this possibility may impart an instrumental urgency to decision-making about imaginary cases. Secondly, even if, for two comparable cases, the actual case carries with it an urgency that the imaginary case does not, there may be other imaginary cases that carry more urgency than either of the two. Third, in many instances we do not know (or even have an opinion about) whether a case is actual or imaginary, and this does not seem to affect our ability to make judgments about the case, or the conviction with which we make them.

Moving to realistic descriptive cases, where the question asked is ‘what would happen in the following circumstances?’ the distinction between actual and merely imaginary cases again seems minimal. Answering the question ‘what would happen if I were to release this actual ball down this actual plane?’ and answering the question ‘what would happen if I were to release some ball with mass x down some plane with slope y ?’ merit exactly the same reasoning process. It might seem that all sorts of things could get in the way of the accuracy of the first answer: the ball might be chipped, or the plane might be warped, or a breeze might be blowing. But this would be to change the case in question; so long as the circumstances described are the same, the reasoning process about the two cases will be the same as well.

But perhaps this is the salient difference: the imaginary case allows idealization, whereas the actual case does not²³. Again, if the issue is that the imaginary case describes an ideal case, whereas the actual case does not, then the cases are not the same; if the imaginary case involves a perfectly smooth plane, whereas the actual case involves one that is only approximately smooth, then the two cases differ, and we are back to our previous failed attempt to distinguish imaginary cases from actual ones²⁴. But we now have the suggestion of a systematic difference, namely, that imaginary cases have the possibility of being described in idealized ways, whereas actual cases do not. But, to the extent that this is what is at issue, then the difference between the cases is not that the one is, as a matter of coincidental fact, actual whereas the other is, as a matter of coincidental fact, imaginary, but rather that the case in question is essentially imaginary: we know without looking at the world that no actual state of affairs corresponds to the case described by the scenario.

One way to characterize this distinction is as follows. The imaginary scenario may be either hypothetical or counterfactual. By “hypothetical” I mean: for all we know, such a situation might indeed obtain; whereas by “counterfactual” I mean: without looking at the world, we know the situation is imaginary²⁵. Then we

²³ For a discussion of the role of idealization in contemporary physics, see the papers collected in Shanks 1998.

²⁴ That is, the question we are concerned with is: what is the difference between a case which is imaginary and a case which is actual, where the cases have exactly the same characteristics.

²⁵ Of course this line is not clean. However, one might think of the distinction as follows: roughly speaking, it is the fact that the world is not effortlessly navigable that leads us to consider cases that are hypothetical, and the fact that it is not maximally replete that leads us to consider cases that are counterfactual.

can say that, to the extent that the contemplation of imaginary exceptional cases raises philosophical issues that the contemplation of actual exceptional cases does not, this is a consequence of the cases being, in the sense that I have just described, counterfactual²⁶.

1.3 THOUGHT EXPERIMENTS

1.3.1 What Is a Thought Experiment?

It may seem remarkable that I have gotten this far into the introduction without mention of thought experiments. The omission is deliberate. The term “thought experiment” has broad and disputed application, and I wish as little of my argument as possible to depend upon one or another delineation of this expression²⁷. Nonetheless, I think any characterization of the term that aims to do justice to the way in which the expression is used will have to

²⁶ I am using these terms in a rather idiosyncratic way. A different distinction, though one not unrelated to the one I have been discussing, can be found in the works of Onora O’Neill. O’Neill suggests that when we are discussing the role of (imaginary) examples in ethics, a distinction can be drawn between examples which are hypothetical : “they consist of more or less specific principles of possible action (e.g. the principle of cheating a gullible customer...) whose moral significance can be determined...by applying the Categorical Imperative; [and] examples [which] are ostensive: they point out acts or persons or lives some of whose features are held to be morally significant (e.g. taking the life of Christ as a model for imitation)” (O’Neill 1986, 7). She continues: “hypothetical examples, being themselves principles of action, must evidently remain indeterminate, even when relatively specific, and so cannot fully determine an act. The acts or persons or lives which are pointed to in ostensive examples may, in themselves, be fully determinate. But their relevance to a case in hand must...be guided by some...understanding of the morally significant aspects of the example...Judgment is therefore always needed...Neither principles nor examples alone can guide action” (O’Neill 1986, 7-8).

²⁷ For a comprehensive list of works discussing thought experiments in science and philosophy, see the “Bibliography on Thought Experiment and Experiment” prepared by David Bzak, which can be found at the end of this volume. At present, the only general book-length treatments of the topic are Häggqvist 1996 and Sorensen 1992a. (In addition, Brown 1991a offers a book-length discussion of thought experiments in the natural sciences.) The papers collected in Horowitz and Massey 1991, and (less directly) DePaul and Ramsey 1998 might also serve as relatively exhaustive introductions to the central issues.

account for at least the following six cases, which I take to be paradigmatic instances of its proper usage. (For those unfamiliar with them, I provide brief summaries of the cases in the Appendix to this chapter):

- Plato’s story of the Ring of Gyges, in which Glaucon tries to bring his listener to see that “those who practice justice do it unwillingly and because they lack the power to do injustice” (cf. Plato c.380BC/1974, 359b).
- Stevinus’s discussion of the inclined plane, in which he seeks to establish that the force required to hold a ball in place along an inclined plane is inversely proportional to the length of the plane (cf. Stevin 1955).
- Einstein’s discussions of moving trains, in which he seeks to establish that a commitment to the view that the speed of light is constant brings with it a commitment to the relativity of simultaneity. (cf. Einstein 1961, 21-27)
- Putnam’s discussion of Twin Earth, in which he seeks to establish that meaning is fixed externally (cf. Putnam 1975a, 223-227).
- Nozick’s discussion of the Experience Machine, in which he seeks to establish that we care about more than hedonistic experience (cf. Nozick 1974, 42-45).
- Thomson’s discussion of the ailing violinist, in which she seeks to establish that abortion may be morally justifiable, even if the fetus is a person (cf. Thomson 1971).

Such practice does run counter to the original use of the term, which was reserved for the contemplation of imaginary cases eliciting physical intuitions²⁸. George Bealer complains:

In recent philosophy there has been an unfortunate blurring of traditional terminology. Rational intuitions about hypothetical cases are often being erroneously called *thought experiments*. This deviates from traditional use, and it blurs an important distinction that we should keep vividly in mind. Traditionally, in a thought experiment one usually elicits a physical intuition (not a rational intuition) about what would happen in a hypothetical situation in which physical, or natural, laws (whatever they happen to be) are held constant but physical conditions are in various other respects nonactual and often highly idealized...A classic example is Newton's thought experiment about a rotating bucket in an otherwise empty space. Would water creep up the side of the bucket (assuming that the physical laws remained unchanged)? Rational intuition is silent about this sort of question. Rational intuitions concern such matters as whether a case is possible (logically or metaphysically), and about whether a concept applies to such cases...[T]o call [such cases] thought experiments is, not only to invite confusion about philosophical method, but to destroy the utility of a once useful term." (Bealer 1998, 207-8).

I think Bealer is right to point out that there is an important distinction to be drawn between rational and physical intuitions, and that there is an accompanying distinction to be made between the sorts of thought experiments conducted in pursuit of philosophical knowledge and those conducted in pursuit of scientific knowledge. Indeed, Section 1.3.3 is devoted to an attempt to classify thought experiments along precisely these grounds. Nonetheless, I do not

²⁸ The term *Gedankenexperiment* is widely thought to have been coined by Ernst Mach in 1883; his paradigmatic example of a thought experiment is the case from Stevinus (described above). (For Mach's use of the term, see Mach 1933/1960 and Mach 1926/1976.) This attribution seems to be mistaken, however. Johannes Witt-Hansen traces the use of the term to the early nineteenth-century Danish Kantian Hans Christian Ørsted, who, according to Witt-Hansen, introduced the term in his 1811 "Prolegomenon to the General Theory of Nature" (cf. Witt-Hansen 1976). (It is uncertain whether Mach knew of Ørsted's use of the term.) (Thanks to Dave Bzdak for uncovering this reference.)

think it misleading to use the term to describe the contemplation of imaginary cases more generally. For I think there are certain structural features that hold in common across all thought experiments.

1.3.2 The Tripartite Structure of Thought Experiments

For each of the examples above, the following characterization captures something important about its fundamental structure:

- (1) An imaginary scenario is described .
- (2) An argument is offered that attempts to establish the correct evaluation of the scenario.
- (3) This evaluation of the imagined scenario is then taken to reveal something about cases beyond the scenario.

So, for example, in the Experience Machine example, the imaginary scenario is the existence of the machine as described, the evaluation of the scenario is that we would consider hooking up to such a machine to be undesirable, and the larger lesson is that "we learn that something matters to us in addition to experience by imagining an experience machine and then realizing that we would not use it" (Nozick 1974, 44). In the Twin Earth example, the imaginary scenario is the existence of a planet on which something qualitatively identical to water has the chemical structure XYZ, the evaluation of the scenario is that speakers of English and Twin-English refer to something different by their use of 'water,' and the larger lesson is that "'meanings' just ain't in the head" (Putnam 1975a, 227). In the Stevinus example, the imaginary scenario is the presence of the balls along the inclined plane connected by the chain at the bottom, the evaluation of the scenario is that the balls will remain in equilibrium, and the larger lesson is that the force necessary to hold a ball in place along an inclined plane varies inversely with the length of the plane. That parallel analyses could be offered for the remaining examples is, I trust, clear²⁹.

²⁹ That characteristic (3) is the intention of thought experiment is strikingly apparent in the following quote from Strawson:

In general, this form of question [that is, questions asked about wildly imaginary cases] may be seen as simply a convenient, if perhaps over-dramatic, way of raising more

Although this positive taxonomy is not especially interesting (its breadth of application comes in part from the imprecision of its categories) its negative counterpart is surprisingly useful. Corresponding to (1), (2) and (3) above are three criticisms:

- (1') Unimaginability: The scenario described is not (fully) imaginable³⁰.
- (2') Unsound Argument: The scenario described is imaginable, but the argument establishing the correct evaluation of the scenario is unsound³¹.
- (3') Inapplicability: The scenario described is imaginable, and the argument establishing the correct evaluation of the scenario is sound, but the conclusion does not reveal about the actual world what the author takes it to reveal.

evidently legitimate types of questions not about hypothetical beings at all, but rather, for instance, about the extent to which, and the ways in which, we might find it possible to reinterpret, within a part of our experience, some of the most general conceptual elements in our handling of experience as a whole...My real concern is with our own scheme, and the models of this chapter are not constructed for the purpose of speculation about what would really happen in certain remote contingencies. Their object is different. They are models against which to test and strengthen our own reflective understanding of our own conceptual structure. Thus we may suppose such and such conditions; we may discuss what conceptual possibilities and requirements they can be seen by us as creating...In all this we need no more claim to be supposing real possibilities than one who, in stricter spheres of reasoning, supposes something self-contradictory and argues validly from it (Strawson 1959, 86).

³⁰ This may be for one of two reasons: the scenario might be incoherent, or it might be underdescribed (either resolvably—such that all we need to do is to fill in some of the details, or irresolvably—such that any filling-in of the details would result in a scenario that is incoherent).

³¹ Again, this might be for one of two reasons. The argument might be independently unsound or it might be situationally unsound. An argument is independently unsound if it is unsound for reasons having nothing to do with the thought-experimental scenario. An argument is situationally unsound if its use depends on appeal to a principle that cannot justifiably be employed in the way that the author wishes. This may be (a) because something about the imaginary scenario implies that the ordinary criteria for application of a particular concept are unavailable or, (b) because the ordinary justification for appeal to a particular principle is missing. (The criticisms I offer in Chapter 4 are criticisms of this sort.)

These criticisms provide a surprisingly fruitful way of classifying criticisms of the sort: “the scenario we are being asked to consider is just too far-fetched.” Those who make unimaginability objections question the very coherence of the exceptional case described³². Those who make unsound argument objections question our ability to use normal modes of reasoning in thinking about the exceptional case³³. And those who make inapplicability objections question the extent to which conclusions drawn on the basis of the exceptional case can be applied (in the way that the author wishes) to normal cases as well. In all three cases, the objection is perhaps best understood as an expression of the legitimate concern that our beliefs, desires, and concepts are deeply tied to our views about which alternate possibilities are salient, so that the imagined disruptions of these patterns of saliency will leave us with too little to base our judgments on. This may be (1) because we cannot fully imagine such a scenario, or (2) because we would not know how to make sense of it, or (3) because we would not know how to apply its lessons to our world³⁴.

³² That is, they contend that the case is either incoherent, or irresolvably underdescribed.

³³ That is, they contend that the argument is situationally unsound.

³⁴ As a test of the fruitfulness of this analysis, consider the sense it makes of the following oft-quoted passage from Kathleen Wilkes. (The passage is cited in Martin 1991 and Oderberg 1993, among others; reference is made to similar passages in Baillie 1993a, Hertzberg 1991, Madell 1991 and Rovane 1993.) Wilkes writes that in thinking about wildly imaginary cases (what I would call clearly counterfactual cases):

either (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 determinately different we do not know what we would want to say about anything (Wilkes 1988, 46).

The taxonomy above allows us to analyze her argument as follows:

³⁵ (2') situationally unsound: “[if] (a) we picture them against the world as we know it...then we picture them against a background that deems them impossible”

So, for instance, someone might object that the idea of a machine that would perfectly simulate experiences is just too far-fetched for us to make sense of, and by that, they might mean one of three things:

- (1') Unimaginability: That we cannot really imagine there being such a machine.
- (2') Unsound Argument: That although we can imagine there being such a machine, if we try to reason about how we would react to the possibility of being hooked up to it, we will inevitably make judgments on the basis of assumptions that would not be relevant under such circumstances.
- (3') Inapplicability: That although we can imagine there being such a machine, and we can make informative judgments about what it would make sense to do and say if we were confronted with the possibility of being hooked up to it, the conclusions that can be drawn on the basis of those judgments do not establish the claim that, in general and in actual cases, “something matters to us in addition to experience.”

This sort of objection is most powerful when it is made against thought experiments which are, in the sense described at the end of section 1.2, counterfactual. But even in the case of hypothetical thought experiments, objections of at least the latter two kinds can be made³⁵. One might (albeit implausibly) object to Stevinus that his evaluation of the scenario he has described is incorrect, or that the case he has described is a special one, from which general conclusions cannot be drawn. Nonetheless, this asymmetry points to a certain distinction among types of thought experiments which this characterization does not capture, and which I describe in the next section.

(3') *inapplicable*: “or (b) we picture them against some quite different background... then we have the realm of fantasy, and fantasy is fine to read; but it does not allow for philosophical conclusions to be drawn”

(1') *irresolvably underdescribed*: “in a world determinately different we do not know what we would want to say about anything”

³⁵ See the detailed discussion of the Einstein-Bohr debate concerning Einstein’s clock-in-a-box thought experiment in Norton 1991, 139-142. See also Allen I. Janis, “Can Thought Experiments Fail?” (Janis 1991).

1.3.3 Three Sorts of Thought Experiments

In the last section, I stressed certain features that I think are common to all thought experiments. These features make it reasonable to describe thought experiment as a mode of reasoning. But there are crucial contrasts to be drawn among types of thought experiment, and in this section, I describe one such difference. What I want to suggest is that there are three basic sorts of questions that might be asked about an imaginary scenario:

- (1) What would happen?
- (2) How, given (1), should we describe what would happen?
- (3) How, given (2), should we evaluate what would happen?

We might call the first type factive (concerning what we think the facts of a situation would be), the second conceptual (concerning what we take to be the proper application of concepts), the third valuational (concerning the proper moral or aesthetic response to a situation). These three questions correspond—very roughly—to the sorts of issues addressed by thought experiments in (1) science; (2) metaphysics and epistemology; and (3) ethics and aesthetics. Of course, this is only a generalization; scientific thought experiments may ask how something ought to be described, epistemological thought experiments may ask what ought to be, and ethical thought experiments may ask what would happen. But although this classificatory scheme rests on exaggerated differences (indeed, my argument for how it is that imagination can lead to new knowledge relies on showing that the line between the first and the second question rests on a false dichotomy), it is still illuminating to help ourselves to this distinction among what is, what is said to be, and what ought to be. So, for instance, we have examples above of each of the three sorts of cases:

- (1) *Factive*: What would happen?
E.g: Einstein, Stevinus
- (2) *Conceptual*: How, given (1), should we describe what would happen?
E.g: Gyges, Twin Earth
- (3) *Valuational*: How, given (2), should we evaluate what would happen?
E.g: Experience Machine, Thomson’s Violinist

Each sort of case raises distinct philosophical puzzles. In the case of factive thought experiments, the issue is: how is it that thinking about something in a new way can lead us to recognize something new about the physical world? (Consider, for example, Stevinus realizing that we already had enough information to figure out what was going on in the case of the inclined plane.) Before we start thinking about the scenario, we do not know which aspects of the situation are going to be relevant. Somehow, however, by imagining something particular, we acquire a sufficiently rich sense of the situation to answer the question. But why does that help? Obviously we are giving ourselves the image, so in some sense we must already know the thing we are trying to find out³⁶. (Where else could the information be coming from, if not from the image we are contemplating?) So the puzzle is: what is enabling us to see this imagined particular as something about which we can make a judgment if we think hard enough? This is the problem that I address in Chapter 2, where I discuss a famous thought experiment of Galileo's.

For conceptual and valuational thought experiments, the issue is: what do we expect to learn about our concepts or values by trying to make sense of this imagined case? Why should thinking about a case that has not occurred or is not going to occur help us understand how we (should) evaluate actual cases? The reasons that we think about imaginary cases are the reasons I discussed in Section 1.2: not every way that things might be is a way things are, and even when things are some way, it may not be so easy for us to reach them and manipulate them. At the same time, one of the ways we are able to learn which features of a particular circumstance are relevant to our evaluation of it is by Mill's methods of comparison, varying the factors which contribute to a situation to see which of them plays which role.

³⁶ No one expects to get rich by paying their right hand with their left! Why then do we expect to gain knowledge by such a technique?

In scientific cases, when we explore physical dependencies, we do (actual) physical experiments. In philosophical cases, when we explore conceptual dependencies, we do thought experiments³⁷. In both, we test various hypotheses by considering cases in which we systematically vary the possible contributing factors. In actual scientific experiments, the circumstances that we modify are circumstances in the world, and the knowledge we gain is empirical knowledge. But in philosophical thought experiments, the circumstances we modify are circumstances that we ourselves have contrived: we conjure up some situation, conjure up modifications of the situation in light of what we wish we knew about the world, and suddenly declare ourselves to have learned something new. But unlike the factive cases, in conceptual and valuational cases we already know what would happen; we are worried here about how we should describe or evaluate it. So the puzzle here is: what could possibly be guiding us in making such a judgment? This is the problem that I address in chapters 3 and 4, in my discussions of the Ship of Theseus (chapter 3) and of personal identity (chapter 4).

With these preliminaries in place, it is time to begin.

³⁷ For related contemporary psychological discussions of conceptual change which have informed my outlook on these questions, see (books) Carey 1985; Keil 1979, 1989; Markman 1989; Wellman 1990; and (anthologies) Margolis and Laurence 1999; Hirschfeld and Gelman 1994; Olson and Torrance 1993; Neisser 1987; Gentner and Stevens 1983.

APPENDIX TO CHAPTER 1

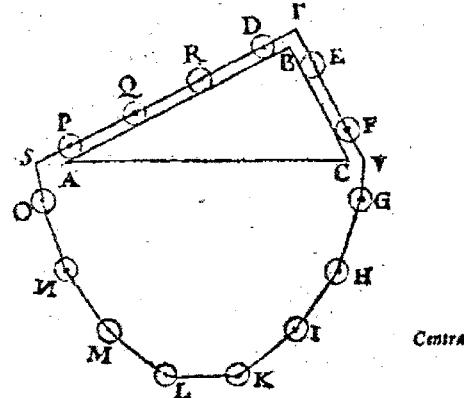
Six Paradigmatic Thought Experiments

(1) *Ring of Gyges*

In the context of a discussion with Socrates concerning the nature of justice, Plato's character Glaucon presents the listener with the following story. Suppose there were a ring that had the capacity to make its wearer invisible, and suppose the ring were given to a person widely considered to be just. Glaucon maintains that no one "would be so incorruptible that he would stay on the path of justice or stay away from other people's property, when he could take whatever he wanted from the marketplace with impunity...Rather, his actions would be in no way different from those of an unjust person." (*Republic*, 359c-360c). Glaucon concludes that "those who practice justice do it unwillingly and because they lack the power to do injustice" (*Republic*, 359b); that is, that the only reason anyone acts justly is out of fear that she will be caught and punished.

(2) *Stevinus's Inclined Plane*

Stevinus considered the question of how much force would be necessary to prevent a ball from sliding down an inclined plane. He imagined a triangular prism on top of which is laid a circular string of fourteen balls:



Introduction

Since assuming otherwise would commit us to the position that perpetual motion is possible, Stevinus concludes that the balls are in a state of equilibrium, that is, that the chain moves neither to the right nor to the left. He next imagines cutting the string at the two lower corners, such that two balls remain along the side with the sharper incline, and four along the side with the shallower incline. Since the balls were in equilibrium prior to the cutting, they remain so afterwards; so the shorter and the longer string of balls are in balance. From this Stevinus concludes that the force required to hold a ball in place along an inclined plane is inversely proportional to the length of the plane. (Stevin, 1955).

(3) *Einstein's Moving Trains*

Einstein asks us to imagine a situation in which there are two groups of people: one group assembled at a certain point along the embankment of a railroad track, and the other group riding on a train that is moving with respect to the embankment. Suppose that lightning strikes the embankment "at two places *A* and *B* far distant from each other" such that those standing on the embankment stand at a point exactly between them. Suppose further that from the perspective of the people standing at that point, the flashes occur simultaneously. Does it make sense to say that the flashes occurred simultaneously, *simpliciter*?

Einstein answers "no." "When we say that the lightning strokes *A* and *B* are simultaneous relative to the embankment, we mean: the rays of light emitted at the places *A* and *B*, where the lightning occurs, meet each other at the mid-point *M* of the length *A*→*B* on the embankment." But the rider of the train "(considered with reference to the railway embankment) is hastening towards the beam of light coming from *B*, whilst he is riding on ahead of the beam of light coming from *A*." So given that the speed of light is constant, the train passenger "will see the beam of light emitted from *B* earlier than he will see that emitted from *A*...We thus arrive at the important result: Events which are simultaneous with reference to the embankment are not simultaneous with respect to the train, and vice versa...unless we are told the reference-body to which the statement of time refers, there is no meaning in a statement of the time of an event." (All quotes taken from Einstein 1961, 21-27.)

(4) Putnam's Twin Earth

Putnam asks his reader to imagine a planet, Twin Earth, which is “exactly like Earth” except that on Twin Earth, “the liquid called ‘water’ is not H₂O but a different chemical whose chemical formula...[will be abbreviated] simply as XYZ.” According to the story, “XYZ is indistinguishable from water at normal temperatures and pressures,” so that when speakers on Twin Earth refer to “water,” they refer to something that gives them exactly the same sorts of experiences (in terms of taste, quenching-of-thirst, wetness, cleansing power, etc.) that water (that is, H₂O) gives to speakers who use the term on Earth. So even though the speakers of English and the speakers of Twin-English use the term “water” in exactly the same circumstances, and even though they may be in exactly the same psychological states as one another when they do so, the extension of the term (that is, what it is that the term picks out in the world) differs in the two cases; “the extension of ‘water’ in the sense of water_E is the set of all wholes consisting of H₂O molecules, or something like that; the extension of water_{TE} is the set of all wholes consisting of XYZ molecules, or something like that.” On the basis of this and other examples, Putnam takes himself to have established that “meanings just ain’t in the head” (Putnam 1975a, 223-227).

(5) Nozick's Experience Machine

Nozick asks his reader to imagine a machine which allows “superduper neurologists” to “stimulate your brain so that you would think and feel that you were writing a great novel, or making a friend, or reading an interesting book” where really “[a]ll the time you would be floating in a tank, with electrodes attached to your brain.” Importantly, from the inside there would be no way to tell that you were ‘hooked-up’: “while in the tank you won’t know that you’re there; you’ll think it’s all actually happening.” Nozick contends that we would consider hooking up to a machine to be undesirable, concluding: “[w]e learn that something matters to us in addition to experience by imagining an experience machine and then realizing that we would not use it.” (See Nozick 1974, 42-45)

(6) Thomson's Ailing Violinist

Thomson tells a story in which while you were asleep, a Society of Music Lovers has you connected to an ailing violinist, so that, for the next nine months, your kidneys will process the fluids he will need for survival. Thomson argues that even if disconnecting him will result in his death, you might still be morally justified in doing so, and she suggests that this, along with other cases, shows that even if we assume that the fetus is a person, abortion might be morally justifiable in some cases. (See Thomson 1971)

CHAPTER 2

Galileo

In this chapter, I discuss one of the most famous thought experiments in the history of science. Through a careful examination of this *factive* thought experiment, I try to show that the guided contemplation of an imaginary scenario can provide us with new scientific knowledge in a way that argument alone cannot.

2.1 ARGUMENTATIVE RECONSTRUCTION

Philosophers who are opposed to all things spooky tend to think that thought experiments in science are (at least in principle) eliminable, and that whatever demonstrative force they have is the result of their being sound arguments dressed up in heuristically appealing clothing. On such a view, a scientific thought experiment's justificatory force comes from the fact that it can be reconstructed as an argument with explicit premises that make no reference to imaginary particulars.

My goal in this chapter is to challenge this view by carefully examining one of the most famous thought experiments in the history of science: that by which Galileo is said to have refuted the Aristotelian theory that heavier bodies fall faster than lighter ones. I will try to show that the thought-experimental format of Galileo's presentation plays an indispensable role in the persuasiveness of his case against the Aristotelian, and that a similar degree of persuasiveness could not be obtained on the grounds of explicit argument alone.

2.1.1 The Elimination Thesis

The view that thought experiments lead to justified conclusions because they are arguments finds clear articulation and powerful defense in a recent pair of papers by John Norton¹. In those papers, Norton puts forth a hypothesis about thought experiments which he calls the *Elimination Thesis*. Paraphrasing slightly, his thesis is this:

The Elimination Thesis: Any conclusion reached by a (successful) scientific thought experiment will also be demonstrable by a non-thought-experimental argument².

As initially formulated, the thesis is ambiguous; it is compatible with both a weaker reading, which I will call the *Dispensability Thesis*, and a stronger reading, which I will call the *Derivativity Thesis*³. In order to formulate these versions, however, I need first to clarify what a number of terms in the thesis mean, and to describe briefly the sorts of arguments which might be offered in its favor.

2.1.2 Clarification of Terminology

A number of terms in the thesis require further elaboration. Let me begin with “thought experiment” and “non-thought-experimental argument.” To draw a conclusion on the basis of a *thought experiment* is to make a judgment about what would happen if the particular state of affairs described in some imaginary scenario were actually to obtain⁴. One might then use that judgment in develop-

¹ Norton 1991 and Norton 1996.

² Cf. Norton 1991, p. 131 and Norton 1996, 336. I should make it clear at the outset that although I am using Norton’s articulation and defense of this thesis as a convenient jumping-off point, the specifics of his position are not my target. For a detailed comparison of my views with Norton’s, cf. 2.5 below.

³ Cf. Norton 1996, 354-8.

⁴ For other recent characterizations of (scientific) thought experiments see, for instance: Bealer 1998; Brown 1986, 1991a, 1991b, 1993a, 1993b, 1995, undated; Carrier 1993; Gooding 1990, 1992, 1993, 1994; Hacking 1993; Humphreys 1993; Irvine 1991; Janis 1991; Kujundzic 1992 and 1995; Laymon 1991; Lipton 1991; Massey 1991, 1995; Nersessian 1984, 1992, 1993; Norton 1991, 1993, 1996;

Continued on next page

ing a more general theory, just as one might use the result of an actual experiment. By contrast, to draw a conclusion on the basis of a *non-thought-experimental argument* is to be led by a process of inductive or deductive reasoning from a set of explicit premises that make no reference to particular hypothetical or counterfactual states of affairs to a correspondingly general conclusion. Again, one might use that conclusion as the basis for endorsing one or another general theory about the phenomena in question. So thought experiments differ from non-thought-experimental arguments in two crucial respects: first, they are not presented as arguments, but rather as invitations to contemplate a way that the world might (have) be(en); and second, they make essential reference to particular hypothetical and counterfactual states of affairs⁵.

What the Elimination Thesis says is that any good scientific thought experiment can be transformed into a non-thought-experimental argument without loss of demonstrative force. Given the characterization just offered, what an elimination will involve is first a process of argumentative reconstruction in which the narrative presentation is replaced by a series of explicit premises sufficient to establish the desired result, and then a process in which those premises that make reference to hypotheticals, counterfactuals, and particulars are replaced by premises in which no such reference is made. If the Elimination Thesis is correct, such a process will preserve completely the thought experiment’s demonstrative force.

What is meant by “demonstrative force”? I will suggest two problematic readings and then one that I will endorse. If the claim that “any conclusion reached by a good thought experiment will also be demonstrable by a non-thought-experimental argument” means no more than that in the reconstruction of a mature science, the conclusions that were (as a matter of fact) reached by thought experiments can be derived from more fundamental principles by means of inference schemes licensed within the science, then the thesis is trivially true. Even if the development of Newtonian

Prudovsky 1989; Shepard 1994; Sorensen 1992a, 1992b; Wilkes 1988; as well as articles collected in Horowitz and Massey 1991, Achinstein and Hannaway 1985, and Shanks 1998. Cf. also Mach 1883/1960, 1897/1976; and discussions in Hesse 1966; Koyré 1960; Kuhn 1964; Popper 1959/1992.

⁵ Cf. Norton 1991, p. 129 and Norton 1996, p. 336.

mechanics relied on a series of crucial thought experiments, its textbook presentation might well establish particular conclusions on the basis of more conventional forms of argument.

On the other hand, there is a reading of “demonstrative force” according to which the thesis is trivially false. If the claim is taken to mean that—as a matter of psychological fact—any conclusion which was reached by a good thought experiment might also have been demonstrated to the person who reached the conclusion by means of a non-thought-experimental argument, then the Elimination Thesis is certainly false. None would doubt the important heuristic and illustrative role played by thought experiments in scientific exploration, and the crucial tasks they play in instruction and informal demonstration.

The proper reading of “demonstrative force” makes the Elimination Thesis epistemologically interesting. On this reading, demonstrative force concerns the role that thought experiments play in living bodies of knowledge: after the moment of discovery and before the end of inquiry. It concerns whether a particular conclusion based on a particular process of reasoning (thought experiment) is thereby *justified*—whether if such a process leads to true beliefs, those beliefs should count as knowledge. So the issue raised by the Elimination Thesis is this: can reasoning about (reasonably) specific entities within the context of an imaginary scenario lead to rationally justified conclusions that—given the same initial information—would not be rationally justifiable on the basis of a straightforward argument? In the next section, I consider two reasons that one might think such eliminations are possible.

2.1.3 The Negative Argument and the Positive Argument

The Elimination Thesis (that is, the thesis that thought experiments are dispensable) can be defended with at least two arguments, one negative, one positive. So, for instance, John Norton argues (negatively) as follows: thought experiments must be arguments because there is nothing else for them to be. “Thought experiments in physics provide or purport to provide us information about the physical world. Since they are *thought* experiments rather than *physical* experiments, this information does not come from the reporting of new physical data. Thus there is only one non-controversial source from which the information can come: it is elicited from information we already have by an identifiable

argument” (Norton 1991, 129). Norton considers this position almost trivial: “the alternative,” he writes, “is to suppose that thought experiments provide some new and even mysterious route to knowledge of the physical world” (Norton 1991, 129)⁶. So the negative argument contends that if we have obtained new information about the empirical world without having obtained new *empirical* information about the empirical world, the only way we could have done so is by means of an argument.

The positive reason that one might think that thought experiments are just arguments in disguise is that “the *analysis* and *appraisal* of a thought experiment will involve reconstructing it explicitly as an argument,” so that “a good thought experiment is a good argument, a bad thought experiment is a bad argument” (Norton 1991, 131; 1996, 335). So the positive argument amounts to saying that if a thought experiment can be *reconstructed* as an argument, then what it was all along *was an argument*⁷. Even if the reason I come to know something by contemplating a thought-experimental scenario doesn’t *seem* to be because there is an argument into which the thought experiment can be reconstructed, it is. The reason my belief is *justified* is because, in the end, the thought experiment *was* a disguised argument all along.

2.1.4 The Dispensability Thesis and the Derivativity Thesis

We are now in a position to recognize that the Elimination Thesis as originally formulated and defended actually involves two distinct claims. These might be stated as follows:

The Dispensability Thesis: Any good scientific thought experiment can be replaced, without loss of demonstrative force, by a non-thought-experimental argument.

⁶ For an endorsement of this alternative, cf. Brown 1991a, 1991b, 1993a, 1993b, 1995.

⁷ “The workings and achievements of any thought experiment can be revealed and captured fully in an explicit argument which employs the same resources” (Norton 1996, 339). Cf. also Norton 1996, 357-8.

The Derivativity Thesis: The justificatory force of any good scientific thought experiment can only be explained by the fact that it can be replaced, without loss of demonstrative force, by a non-thought-experimental argument.

Loosely put, the Dispensability Thesis says that we can always get from here to there without appeal to a thought experiment. If a thought experiment legitimately transports us from one state of belief to another, a non-thought-experimental argument could too. Thought experiments may be convenient and efficient ways of reaching conclusions about the physical world, but they have only the advantage that a car has over walking; they get us where we want to go much more quickly, but they don't get us anywhere we couldn't reach by more pedestrian means.

The Derivativity Thesis says that not only can any good scientific thought experiment be replaced, without loss of demonstrative force, by a non-thought-experimental argument, but that to the extent that a good scientific thought experiment has demonstrative force, it is *because*, deep down, the thought experiment *is* an argument. We may be misled by the surface features of the case to think that something non-argumentational is doing justificatory work, but we are wrong. The *reason* the Dispensability Thesis is true is that all that was *ever* justificatorily at play was something argumentative. What looked like a car turned out to be propelled by foot power all along (like a child's go-car, or one of the vehicles on *The Flintstones*). So the Dispensability Thesis says we *can* get by without what we commonly call thought experiments as such; the Derivativity Thesis tells us that we already *do*.

In the next section, I challenge the Dispensability Thesis by showing that it does not hold true of a widely-acclaimed thought experiment of Galileo's. I choose this case for two reasons. First, since this particular example is generally treated as the paradigm of an effective thought experiment, diagnosing the source of its success is itself a worthwhile endeavor⁸. Second, challenging the

⁸ For some of the many discussions of this and related thought experiments of Galileo, cf. Brown 1986, 1991a, 1991b, 1993a, 1993b, 1995; Cargile 1987; Clement 1983; Koyré 1939, 1960; Kuhn 1964; Norton 1996; Prudovsky 1989; Sorensen 1992a. In my discussion below, I follow the somewhat unfortunate practice of considering this thought experiment outside of both its historical and textual contexts. As a partial remedy to this misleading presentation, I refer the reader to some of the many general discussions of Galileo's work and its context; one might fruitfully begin with: Butts and Pitt 1978; Claggett 1959; Clavelin 1974; Cooper 1935; Damerow *et al.* 1992; Dijksterhuis 1961; Drake 1977, 1989, 1990. See also sources listed in the next three footnotes, and in the Galileo bibliography at the end of this volume.

Continued on next page

Dispensability Thesis in this way allows me to shed light on the Derivativity Thesis as well. Obviously, if the Dispensability Thesis is false, the Derivativity Thesis is too; the more interesting question is whether some alternative explanation can be offered of the thought experiment's success. I try to say something positive about this question in Section 4.

2.2 GALILEO'S THOUGHT EXPERIMENT AND ITS RECONSTRUCTION

2.2.1 Galileo's Thought Experiment

Perhaps the most famous thought experiment in the history of western science is the thought experiment with which Galileo is credited with having refuted the Aristotelian view that the speed with which a body falls is directly proportional to its weight⁹. The thought experiment appears in his last and most mature work, the *Discourse Concerning Two New Sciences*, in the context of a more general discussion of the possibility and nature of motion in a void¹⁰. Galileo's goal in the section as a whole is to establish that "if one were to remove entirely the resistance of the medium, all

⁹ Challenges to the Aristotelian thesis—both empirical and conceptual—had appeared in a number of mid- and late-sixteenth century works. (For relevant passages, see: [Cardan] Cooper 1935, 7-77; Damerow *et al.* 1992, 365; [Tartaglia] Drake and Drabkin 1969, 63-143, esp. 120ff; Damerow *et al.* 1992, 378; [Benedetti] Drake and Drabkin 1969, 147-237, esp. 206; Drake and Drabkin 1969, 31-41; Dijksterhuis 1961, 269-71; Drake 1989, 27-30; [Stevin] Cooper 1935, 77-80; Dijksterhuis 1961, 324-9.)

¹⁰ Galileo himself had produced a less conclusive version of the famous thought experiment as early as 1590 in an unfinished and unpublished dialogue on motion; cf. Cooper 1934, 80-90; Drake and Drabkin 1969, 331-77; Galileo 1590/1600, 26-38, esp. 29 (National Edition 265). For an interesting discussion of whether Galileo ever actually performed such an experiment, cf. Cooper 1935; Drake 1977, 1989, 1990; Drake and Drabkin 1969; Koyré 1960; Segre 1989.

Continued on next page

materials would descend with equal speed" (Galileo 1638/1989, p. 116)¹¹; the thought experiment in question leads to the weaker conclusion that "both great and small bodies, *of the same material*, are moved with like speeds" (Galileo 1638/1989, p. 109, italics added, bracketed word replaced).

The view that Galileo is challenging is that "moveables differing in heaviness are moved in the same medium with unequal speeds, which maintain to one another the same ratio as their weights *gravità*" (Galileo 1638/1989, p. 106). That is, he is challenging the view that heavier bodies fall faster than lighter ones, and that they do so in direct proportion to their heaviness. On the version Galileo takes himself to be opposing, the proportionality is linear; "a moveable ten times as heavy as another, is moved ten times as fast" as the other (Galileo 1638/1989, p. 106)¹².

The famous thought experiment, rephrased slightly, is the following. Imagine that a heavy and a light body are strapped together and dropped from a significant height¹³. What would the Aristotelian expect to be the natural speed of their combination? On the one hand, the lighter body should slow down the heavier one while the heavier body speeds up the lighter one, so their combination should fall with a speed that lies between the natural speeds of its components. (That is, if the heavy body falls at a rate of 8, and the light body at a rate of 4, then their combination should fall at a rate between the two (cf. Galileo 1638/1989, p. 107).) On the other hand, since the weight of the two bodies com-

Since my primary purpose in this paper is not historical, I will focus only on Galileo's 1638 presentation of the refutation, bracketing the interesting question (a question not without philosophical import) of why it was that such a simple and obvious mistake remained part of the West's scientific world view for nearly 2000 years.

¹¹ Here and elsewhere, I have made use of Drake's 1974/1989 translation of the *Discorsi*. For easy cross-referencing with the more widely available (though less reliable) Crew and De Salvio translation, page references are to the National Edition (except where noted).

¹² Cf. Aristotle *Physics* 215a24-216a21; *On the Heavens (De Caelo)* 301b (Ackrill 1987).

¹³ Note that in the remarks that follow, all references to bodies should be understood as referring to bodies of the same material. For the purposes of my discussion, this constraint is irrelevant.

bined is greater than the weight of the heavy body alone, their combination should fall with a natural speed greater than that of the heavy body. (That is, if the heavy body falls at a rate of 8 and the light body with a rate of 4, their combination should fall at a rate greater than 8.) But then the combined body is predicted to fall both more quickly, and more slowly, than the heavy body alone (cf. Galileo 1638/1989, 107-8). The way out of this paradox is to assume that the natural speed with which a body falls is independent of its weight: "both great and small bodies...are moved with like speeds" (Galileo 1638/1989, 109).

2.2.2 Reconstruction of the Galileo Case

Transformed into an argument that conforms to the strictures of the Elimination Thesis, Galileo's reasoning can be reconstructed as follows¹⁴. The first claim of the Aristotelian is that:

(1) Natural speed is mediative.

That is, natural speed is a property such that if a body A has natural speed s_1 , and a body B has natural speed s_2 , the natural speed of the combined body A-B will fall between s_1 and s_2 .

The second premise of the reconstruction is that:

(2) Weight is additive.

That is, weight is a property such that if body A has weight w_1 , and body B has weight w_2 , the weight of the combined body A-B will be equal to the sum of w_1 and w_2 .

From these two premises (plus the assumption that not all weights and natural speeds are either zero or infinite), it follows that:

(3) Natural speed is not directly proportional to weight.

For the first is a mediative property, whereas the second is an additive property, and a mediative property cannot be directly proportional to one that is additive. Furthermore, the only way to

¹⁴ For Norton's reconstruction of this case, cf. Norton 1996, 340-5; I discuss his version in 3.4 below.

maintain (1), (2) and (3) simultaneously is to assume that all natural speeds are the same. Then weight might be additive and natural speed (in a vacuous sense) mediative, with no contradiction thereby implied. Thus natural speed is shown to be independent of weight.

2.2.3 Four Ways out for the Aristotelian

If the Dispensability Thesis is true, then Galileo's thought experiment should be replaceable by some non-thought-experimental argument without loss of demonstrative force. My goal in the next two sections (2.2.3 and 2.2.4) is to show that the reconstruction presented in section 2.2.2 is not such an argument¹⁵.

I begin my case by pointing out that there are a number of "ways out" for the defender of the view that natural speed is directly correlated with weight—a view which, for the sake of convenience, I will call the Aristotelian view. These ways out involve denying premises (1) and (2) by proposing a series of alternative hypotheses about the physical properties of strapped-bodies; that is, bodies of the sort described by the thought experiment. The point of talking about these ways out is to show that there are ways to maintain the negation of (3) by adopting alternatives to (1) and (2), and adopting these may well be less disruptive to the Aristotelian picture than giving up (3). What I will suggest below is that these ways out, though logically available, run counter to certain tacit knowledge about the physical world. It is for this reason that, when the case is presented as a thought experiment, they do not even occur to us. To block them as moves in a straight argument, however, requires metaphysical commitments that seem not to be at play in the thought experiment itself. What these commitments are, and what role I think they *actually* play in Galileo's reasoning is a point I will turn to after presenting the four ways out.

The first ways out would be for the Aristotelian to deny that the properties in question are *determinate* for strapped-together bodies in one of the following two ways.

¹⁵ Below I suggest reasons for thinking that this will be true for *any* argumentative reconstruction that conforms to the strictures of the elimination thesis.

(4) Natural speed is not physically determinate for strapped-bodies¹⁶.

(5) Weight is not physically determinate for strapped-bodies.

That is, she might reject (1) or (2) on the grounds that they presuppose that natural speed and weight are properties that apply universally, even to bodies that are in some way monstrous¹⁷. Since strapped-bodies are odd entities, she might say, they need not be governed by the sorts of laws that govern ordinary objects. In particular, they need not have determinate natural speeds or weights.

The third way out for the Aristotelian would be to avoid the conflict between (1) and (2) by saying that there *is* a fact of the matter about whether a strapped-body is one body or two, and that its physical properties in falling will depend on the answer to this question. She might say:

(6) Natural speed and weight are mediative for strapped-bodies that are united. Natural speed and weight are additive for strapped-bodies that are unified.

That is, sometimes when two bodies are strapped together, they are merely *united* and remain, as a matter of fact, two objects; sometimes, when they are strapped together, they are *unified* and form, as a matter of fact, a single object. In the first case, both weight and speed will be mediative; so that the combined body will have a weight intermediate between those of the two original bodies, and fall with a natural speed that lies between the two original speeds. In the second case, both properties will be additive; so that the weight of the unified body will be equal to the sum of the weights of its component parts, and its natural speed correspondingly equal to the natural speeds of the two combined. Since the mediativity of the properties holds only with respect to united pairs of objects, and the additivity only with respect to unified single objects, there is no way that the Aristotelian can be forced to a

¹⁶ For a version of this "way out," cf. Koyré 1968, 51.

¹⁷ Galileo preemptively deals with this by getting a concession from Simplicio straight away that "for every heavy falling body there us a speed determined by nature such that this cannot be increased or diminished except by using force or opposing some impediment to it" (Galileo 1638/1989, 107); cf. also Drake's footnote 40 in Galileo 1638/1989 at (modern pagination) 66.

contradiction. What she is forced to accept, however, are radical discontinuities in nature. A body, united, might be falling steadily at a rate of, say, six, and suddenly, should its parts happen to become unified, begin falling at a rate of, say, twelve.

But the Aristotelian can avoid the problem of discontinuity. A fourth way would be for her to say that given two bodies that fall together there is a fact of the matter about their degree of connectedness, and that this determines their physical properties when falling. The claim would be:

- (7) Natural speed and weight for strapped-bodies are determined by a degree of connectedness (C) such that the speed/weight of B_1 -strapped-to- B_2 where B_1 has w_1 and B_2 has w_2 will be: $(C)(w_1+w_2) + (1-C)((w_1+w_2)/2)$ ¹⁸.

We let C measure the degree of connectedness between the two bodies; that is, we let it be a number between zero and one that corresponds to the degree to which two bodies that fall together are unified: if the bodies are completely unified, C will take a value of one; if the bodies are completely disunified (that is, united), C will take a value of zero. For intermediate cases, the value will be between these two, and the speed and weight of the combined body will lie between the mean and the sum of the two initial values. So if the two bodies are completely unified, the additive law will apply completely; if the bodies are merely unified, the mediative law will apply throughout; and for intermediate cases, some proportional average will be found between them. Thus the assumption that natural speed is a function of weight can be maintained, and it can be maintained without violation of continuity. How? Under the assumption that degree of connectedness is a relevant physical property. That this way out too seems not to be a

¹⁸ To keep the equation minimally complicated, I have made a number of trivial simplifying assumptions. I have assumed that the units for measuring weight and natural speed correspond so that the number representing an object's weight is the same as the number representing its natural speed; and I have assumed that the natural speed of two merely unified bodies is the mean of their individual natural speeds.

live option brings us to the point where I will suggest my alternative explanation of what is going on.

2.2.4 What the Reconstruction Misses

I want to begin by thinking about ways in which the four ways-out might be blocked. And I want to start with the most obvious: appeal to two broad, defeasible, tacit assumptions, each of which captures an important feature of our representation of experienced reality. One is that, for any body that one might encounter, there is a determinate fact concerning its weight and natural speed. That is:

- (8) Natural speed and weight are physically determined. The other is that there is *no* determinate fact whether strapped-bodies are one object or two. That is:
- (9) Entification is not physically determined.

What (8) says is that a particular question about natural properties has a determinate answer. Any body, no matter how oddly shaped, will have a particular weight and a particular natural speed that are fixed by the world. What (9) says is that a particular question of entification has an indeterminate answer. Whether we consider a strapped-together body to be a single object, or two objects held together by a strap, or indefinitely many objects held together by internal forces, is merely a question of the aspect under which we choose to view that object. The answer to the question "how many objects?" does not follow from any *physical* property we might discover; it is a question about our words, not a question about the world.

These two premises are sufficient to eliminate the "ways out" enumerated above. If (8) holds, then (4) and (5) are not available as lines of escape; if (9) holds, then neither (6) nor (7) can be appealed to as a means of avoiding the Galilean conclusion. And the *way* in which they eliminate (4)-(7) is very different from the way that a simple reassertion of (1) and (2) would. They show not only *that* there is something wrong with (4)-(7) as descriptions of the way the world is, but *why* there is something wrong with them. They show *what* about our tacit understanding of physical reality, and of our instincts concerning plausible candidates for physically relevant and irrelevant properties, is missed by someone who appeals to (4) or (5) or (6) or even (7).

What this reveals is that the initial reconstruction of the Galilean thought experiment (as presented in section 2.2.2) fails to capture what is really doing the work in the case. As hypotheses about the ways strapped-bodies might behave in fall, (4), (5), (6) and especially (7) are in principle available as alternatives to (1) and (2); just as natural speed might be mediative and weight additive, it might be that both natural speed and weight depend on the degree to which the two bodies are connected. So if (1)-(3) were truly capturing what is going on in the Galilean thought experiment, there would be ways out for the Aristotelian that would allow her at least to shift the burden of proof back to the Galilean.

That these ways out do not seem available when the thought experiment is presented in its unreconstructed form shows that this eliminative reconstruction has failed to capture its original demonstrative force. (What has been lost is the way in which, by evoking tacit knowledge about the how falling bodies actually behave, the thought experiment preemptively precludes such ways out.) Accordingly, I tried to come up with a reconstruction sufficiently strong to rule out (4)-(7) in a similarly categorical and decisive manner. This involved appeal to two rather comprehensive and metaphysical-sounding principles, namely (8) and (9). (8) and (9) give background support to (1) and (2) and thereby help to establish (3).

But just as (1) and (2) are too weak to capture the way in which alternative hypotheses concerning fall are ruled out by the thought experiment, (8) and (9) are too strong. They represent approximate articulations of defeasible assumptions about the physical world. But as they stand, they articulate principles that have less certainty than the conclusion they are taken to support. Prior to contemplation of the case Galileo describes, the Aristotelian may be committed to *something* like (8) and *something* like (9), but he is certainly not committed to them unmodified. To reconstruct the case as an explicit argument with some version of (8) and (9) among the premises would require enumerating outright their defeasibility conditions. But this is something he does not know how to do. Contemplation of the case Galileo describes *brings him to see* that these principles are not defeated in *this* case. And it is this recognition that serves as the basis for the case's power. No austere argumentative reconstruction will be able to do this,

because part of the thought experiment's function is to bring the Aristotelian to accept certain *premises*. In the next section, I will discuss what makes belief in these premises *new*, and what makes it *justified*.

2.3 DENYING THE DISPENSABILITY AND DERIVATIVITY THESES

2.3.1 Rejecting Reconstruction: What the Thought Experiment Does

If the Dispensability Thesis were correct, then the conclusion established by Galileo's thought experiment—that “both great and small bodies...are moved with like speeds” (Galileo 1638/1989, 109)—*should* be demonstrable by means of a non-thought experimental argument. “Demonstrable” here means: rationally justifiable on the basis of the same background conditions. So let us spell out what the background conditions in question are.

For the Aristotelian, daily experience seems to confirm the theory that heavier bodies fall faster than lighter ones. Just as the Galilean sees cases where lighter bodies fall more slowly than heavy ones as exceptional, so the Aristotelian sees as crying out for explanation those cases where the rate of fall is (nearly) simultaneous. That is, when gold beaten into a very thin leaf reaches the ground more slowly than a solid lump of the same material, the Galilean must posit some factor that explains the divergence of this result from the generally expected outcome¹⁹. Similarly, when the Aristotelian sees two stones of very different weights fall to the ground with like speeds, the circumstance requires diagnosis and explanation.

So far all I have pointed out is the possibility of maintaining theoretical commitments in the face of apparent counter-evidence. This can be done by appealing to additional principles that explain away anomalous data by showing that the phenomena at issue are

¹⁹ Cf. Galileo 1638/1989, 109.

subject to the fundamental principle in question, but that the world's complexity has prevented them from manifesting this. So the Galilean might appeal to air resistance, the Aristotelian to the fact that the bodies have not been dropped from a height sufficiently great²⁰. Such explaining-away of recalcitrant exceptions is not a desperate move by a failing paradigm; it is a fundamental element of doing normal science in a non-ideal world²¹.

The point of this discussion is to give a better sense of the background conditions under which the argumentative reconstruction must be demonstratively forceful if the Dispensability Thesis is to be shown to hold in this case. The thesis tells us that some non-thought experimental reconstruction of the case Galileo presents should be able to do the same thing that the thought-experimental version does: lead the Aristotelian *from the same background assumptions to the same rationally justified conclusions*. The standard reconstruction presented in Section 2.2.2 fails in this regard, as does the strengthened version presented in Section 2.2.4. And, I contend, *any* argument satisfying the Elimination Thesis is likely fail in the same way.

Why? Because prior to the thought experiment, the Aristotelian is explicitly committed to the *negation* of (3), and this background commitment serves as a filter through which apparently contrary evidence will inevitably be reinterpreted. Any argu-

²⁰ In Galileo's dialogue, when Salviati points out that a rock of two pounds and a rock of twenty pounds will strike the ground nearly simultaneously when dropped from a height of one- or two-hundred feet, Simplicio retorts that this may just be a result of not having given the objects enough falling-time for the differences to become apparent (see Galileo 1638/1989, 109-10). Simplicio says: "Perhaps from very great heights, of thousands of braccia discrepancies would follow which are not seen at these lesser heights." That is, Simplicio suggests that perhaps the reason we do not observe the sorts of differences which Aristotle's theory predicts is that we are making observations under non-idealized circumstances, and that were we to eliminate distorting elements—like the fact that the fall is only a few hundred feet—the true phenomena would reveal themselves. Galileo provides Salviati with a rather sharp retort. (Galileo 1638/1989, 110). See also (pagination here refers to the modern text) Galileo 1638/1989, footnote 44 at 69.

²¹ Of course, it is a commonplace in the history of science that at a certain point the burden of positing literal or metaphoric epicycles becomes too great, and the theory—especially when there is a simpler alternative available—collapses under the weight of its own internal complexity. But it is a similarly well-established commonplace in philosophy that theories are underdetermined by evidence, and that extra-scientific considerations do some of the work in determining theory choice. (Cf. Kuhn 1970.)

ment which satisfies the Elimination Thesis (that is, any argument with explicit premises which make no reference to particular hypothetical or counterfactual states of affairs) can be reframed by the Aristotelian as a *reductio*. (1) or (2) or (8) or (9) will simply be denied, once their implications are made evident. What is remarkable about the thought-experimental presentation is that it is able to undermine this framework-shaping assumption.

So I conclude, tentatively, that the Dispensability Thesis (and *a fortiori* the Derivativity Thesis) is false. Suppose, however, that somehow it were possible to come up with an argumentative reconstruction that almost exactly captures the strength and limits I have attributed to Galileo's thought experiment²². Would we then be justified in accepting the deeper methodological claim put forth in the Derivativity Thesis: that the *justificatory force* of whatever beliefs I hold via thought experiment is a function of the thought experiment's argumentational essence? In the next section (2.3.2), I offer reasons for thinking that the answer to this question is "no."

2.3.2 Rejecting the Positive Argument: What Makes these Beliefs New?

The Derivativity Thesis says that the justificatory force of any good scientific thought experiment can only be explained by the fact that it can be replaced, without loss of demonstrative force, by a non-thought-experimental argument. So rejection of the thesis can take two forms: denying that the justificatory force of a particular scientific thought experiment *can* be explained this way, and denying that it *can only* be explained this way. The simplest way of doing the former, of course, would be to show that the Dispensability Thesis is not true of some thought experiment; obviously, if there is no non-thought-experimental argument with which the thought experiment can be replaced without loss of demonstrative force, than the justificatory force of the thought experiment cannot be explained by the possibility of such replacement. But for the sake of argument, I am supposing that the

²² Suppose, for instance, we were to enumerate the defeasibility conditions of (8) and (9), and we included these modified versions as premises alongside (1) and (2).

Dispensability Thesis is (at least approximately) true. This leaves two avenues for denying the Derivativity Thesis: denying that the argumentative reconstruction explains the justificatory force of some thought experiment *at all*, and denying that it explains such justificatory force *entirely*.

As I discussed above (2.1.3), two sorts of defense are offered for the Derivativity Thesis. The negative argument contends that thought experiments are arguments because there is nothing else for them to *be*; the positive argument contends that thought experiments are arguments because their “analysis and appraisal” involves explicit argumentative reconstruction. In this section I will address the positive thesis, suggesting that it fails to get at what is most interesting about thought experimental reasoning; in the final section, I offer some thoughts about what sorts of alternative justifications are available such that the negative thesis too is untenable.

It is a mistake, I contend, to think that the *reason* conclusions drawn from thought experiments are justified is because thought experiments have argumentational analogues (if indeed they do). Rather, I want to suggest that the Aristotelian comes to have novel justified true beliefs about the empirical world not because he has (whether he knows it or not) followed along the path of a recognized argument form, but rather because he has performed an act of introspection that brings to light heretofore unarticulated and (because he lacked a theoretical framework in which to make sense of them) heretofore implausible tacit beliefs. There are two things that I need to show myself able to explain. The first is how it is that knowledge has been *gained*. In what way is it that the Aristotelian has come to believe something *new*? The second is how it is that *knowledge* has been gained. In what way is it that the Aristotelian has come to believe something *justified*? I will answer the first in the remainder of this section, and the second in Section 2.3.3.

In addressing the issue of novelty, I will begin with brief remarks about what I think is *not* at issue, and then say something about where I think the important questions rest. One might say that the beliefs in question are not new since, in some sense, the Aristotelian had access to them before. After all, he has acquired no new knowledge of the external world; all he has done is reshuffle tacit beliefs he already held, coming to see their implications. But this view of what makes knowledge new is surely too stringent;

it would, among other things, rule out all mathematical reasoning as a potential source of new knowledge. On the other hand, it seems too weak to say that a belief is new if it merely results from putting together two explicitly held beliefs that have not, for the individual in question, been previously connected. If I believe that snow is white and I believe that crows are black, but I have never thought about the two at the same time, it seems wrong to say that the belief that snow-is-white-and-crows-are-black should count as a *new* belief for me²³. In any case, without spelling out precisely what it is that makes some beliefs new and others mere implications, there is a simple reason to think the Aristotelian’s belief that the speed at which a body falls is independent of its weight should count as a new belief for him: until recently he was explicitly committed to the truth of its negation. This alone suggests that—whatever the implicative relation between his prior commitments and this view about natural speed—the belief should count as new.

But there is a deeper and more interesting way that the belief is new, and that is the following. The thought experiment that Galileo presents leads the Aristotelian to a reconfiguration of his conceptual commitments of a kind that lets him see familiar phenomena in a novel way. What the Galilean does is provide the Aristotelian with conceptual space for a new notion of the *kind of thing* natural speed might be: an independently ascertainable constant rather than a function of something more primitive (that is, rather than as a function of weight). It is in this way, by allowing the Aristotelian to make sense of a previously incomprehensible concept, that the thought experiment has led him to a belief that is properly taken as *new*.

What this suggests is that the Derivativity Thesis is *missing the point* of what makes the Galileo case work as it does. The recog-

²³ Except under very odd circumstances. Suppose I believe that crows are black because I live in a village where there are crows, and I have seen many of them. I also believe that snow is white, because I have read about it in books. In my village, it is taboo to think about black things and white things at the same time, because this is thought to allow the evil spirit access to the soul. One day, I leave my village for the north, and I observe a crow circling above a field of snow. I find the image aesthetically striking, and I ask myself why this is so. In analyzing my response to the visual experience I am having, I realize with a start: “Snow is white and crows are black.” In such a case, it seems plausible to suggest that this is a new belief.

nition that natural speed is independent of weight comes not from tracing the implications of antecedent commitments to (1) and (2), which, after all, lead to the denial of a position to which the Aristotelian is explicitly committed (and thence to retreats such as the four ways out). The recognition comes from the sudden realization, on the part of the Aristotelian, of the conceptual possibility of a certain sort of physical property. Prior to contemplation of the case, there was no room on the Aristotelian picture for the thought that natural speed might be constant, not varying—that it might be dependent not on some specific features of the body in question, but only on the fact that it is a body at all²⁴. After contemplation of the case, there seems to be no conceptual space for the view that it might be variable.

Tempting as it might be to digress, this is not the place for a general discussion of incommensurability across theory-change. What is important for our purposes is only this: one of the things that enables this rather striking shift in the representation of physical reality is that the Aristotelian recognizes that there are experientially possible objects—strapped-together bodies—for which the defeasibility conditions of (8) and (9) are not met (that is, objects of which (8) and (9) hold true), and that these are objects for which his old notion of natural speed simply *does not make sense*. If entification is arbitrary and natural speed and weight are fixed by the world, then a feature-dependent notion of natural speed is just plain incoherent. So one way of thinking about how the thought experiment works is this: it brings the Aristotelian to recognize the inadequacy of his conceptual framework for dealing with phenomena which—through the contemplation of this imag-

²⁴ To get a sense of how odd this transition is, try thinking about *weight* as something dependent not on the specifics of the body in question, but as something constant for all bodies. (That is, to ascertain a body's weight, we would not need to know anything more about it than the simple fact that it is a body.) Clearly this would be a major conceptual readjustment; one might even be inclined to say that we aren't talking about *weight* anymore, since whatever sort of thing weight is, it is surely something that depends upon specific features of the bodies to which it applies. The analogy is not perfect, since part of what happens as a result of thinking about the Galileo case is that it becomes apparent that there is no physical application for the Aristotelian idea of natural speed; like phlogiston, it disappears into the ether of abandoned concepts.

inary case—he comes to recognize as always having been part of his world.

What this suggests is that “the *analysis* and *appraisal* of a thought experiment” need not “involve reconstructing it explicitly as an argument,” in such a way that “a good thought experiment is a good argument, a bad thought experiment is a bad argument” (Norton 1991, p. 129; 1996, 335). After all, the argument from (1) and (2) to (3) is no better or worse than the argument from (not-3) to (not-1) or (not-2). Like an experiment, *part* of what makes a thought experiment good or bad is the validity of the procedure by which the same result can be repeatedly obtained. But another thing that distinguishes good thought experiments from bad is their ability to direct the reader’s attention to inadequacies in her conceptual scheme that she herself recognizes immediately, as soon as they are pointed out to her. It is *this*, I want to suggest, that grounds her new beliefs. Of course, I have said nothing so far about what might make these beliefs *justified*. It is to this issue that I turn in the Section 3.3.

2.3.3 Rejecting the Negative Argument: What Makes these Beliefs Knowledge?

Thought experiments work in a variety of ways. By describing appropriately selected imaginary scenarios, they provide contexts within which sense can be made of previously incomprehensible conceptual distinctions²⁵. This happens when two features that are constantly conjoined in our representations of all actual cases are imaginatively separated in the thought-experimental scenario in a way that shows them to have been isolatable all along. And by describing specific situations, thought experiments, like analogical reasoning in general, can justify conclusions about particular cases without explicit or implicit appeal to more general absolute principles²⁶. Many of the higher-level principles by which we negotiate

²⁵ By this I mean they do something like what the answer to a riddle does in making suddenly intelligible what previously appeared to be a nonsensical description. Cf. Cavell 1979, 156–7.

²⁶ For an articulation and defense of such a view of analogical reasoning, cf. Sunstein 1993 and 1996.

the world are *defeasible*, and the determination of their applicability to particular situations must be made on a case-by-case basis. By bringing the reader to focus on particulars, thought experiments can help the reader distinguish warranted from unwarranted applications of the principle in question²⁷.

So far, however, this has little to do with *justification*. After all, as Norton would ask, if the thought experiment is not an argument, why should we put faith in its conclusion? What I want to explore in this closing section is one possible answer: that thought experiments rely on a certain sort of *constructive participation* on the part of the reader, and that the justificatory force of the thought experiment actually comes from the fact that it calls upon the reader to perform what I will call an *experiment-in-thought*.

An experiment-in-thought is an *actual* experiment; the person conducting the experiment asks herself: “What would I say/judge/expect were I to encounter circumstances XYZ?” and then *finds out* the (apparent) answer. This technique is common in linguistics, where the methodology is used to ascertain the grammaticality of sentences, the meanings of phrases, the taxonomic categories of words, and so on²⁸. And it is, on one view at least, a central element of moral reasoning: we think about particular imaginary cases, observe the judgments that they evoke in us, and use these judgments as fixed points in developing our moral theories²⁹.

How does this connect with the Galileo case? What kind of experiment-in-thought plays a role there? Answer: by thinking about the case in question, we discover what sorts of motions and objects we think are possible in the world. Do we think objects can be strapped together? Yes, we do. Do we think objects fall with radical discontinuities in speed? No, we think they do not. Do we think entification is something that is fixed by the world? No,

²⁷ This often happens when the particulars are sufficiently well-sketched to evoke practical as well as theoretical responses. Cf. also the literature on mental modeling, references to which can be found in Nersessian 1993, as well as at Thomas undated.

²⁸ Cf. Thomason 1991, 247: “When linguists want to test hypotheses about the structure of a particular language, their methodology crucially involves thought experiments in a...literal sense: real experiments carried out in thinking.”

²⁹ Cf., to choose an example nearly at random, consider Thomson 1986, 257: “it is...our moral views about examples, stories, and cases which constitute...data for moral theorizing.” I do not intend to be taking a stand here on questions of moral epistemology.

we do not. Do we think weight and natural speed are fixed by the world? Yes, we do. We *come to recognize* that we have these beliefs by contemplating the imaginary case in question; thinking about the case is what brings us to the realization that we believe what we do. *And*—and this is where the justificatory work comes in—the fact that we have these beliefs gives us *prima facie* warrant to think that they are true.

But why? Why should we think that our pre-theoretical beliefs about the structure of the physical world are reliable? In *The Science of Mechanics*, a few pages after introducing the expression “*Gedankenexperiment*”, Mach writes:

Everything which we observe imprints itself uncomprehended and unanalyzed in our percepts and ideas, which then, in their turn, mimic the process of nature in their most general and most striking features. In these accumulated experiences we possess a treasure-store which is ever close at hand, and of which only the smallest portion is embodied in clear articulate thought. The circumstance that it is far easier to resort to these experiences than it is to nature herself, and that they are, notwithstanding this, free, in the sense indicated, from all subjectivity, invests them with high value (Mach 1883/1960, 36).

So one possible explanation is the one Mach gives. We have stores of unarticulated knowledge of the world which is not organized under any theoretical framework. Argument will not give us access to that knowledge, because the knowledge is not propositionally available. Framed properly, however, a thought experiment can tap into it, and—much like an ordinary experiment—allow us to make use of information about the world which was, in some sense, there all along, if only we had known how to systematize it into patterns of which we are able to make sense³⁰.

This, of course, is the beginning not the end of an answer to the question. But it is sufficient for the modest aim of this chapter. What I have been trying to show is that *something* besides argument might give justificatory force to thought experimental reasoning. The alternative I have proposed has been this: By focusing on imaginary scenarios and making reference to particulars, thought experiments can provide a fulcrum for the reorganization of conceptual commitments; this explains the way in which they

³⁰ For further discussion of these themes, cf. Kuhn 1964.

can provide us with novel *information* without empirical input. And by bringing the reader to perform experiments in thought, thought experiments can lead us to reject shaky (and ultimately false) theoretical commitments in light of newly systematized but previously inarticulable *knowledge* about the way the world is.

The justificatory force of thought experiments is thus parasitic on the extent to which the messy twisted web of background beliefs that underpin our navigation of the world are rightly considered knowledge. To establish this, on coherentist or evolutionary or empiricist grounds, would be an enormous undertaking, and one which I will not even begin to endeavor here³¹. But I hope I have given you some sense, at least, of why I am not convinced that even if the Dispensability Thesis is true, the Derivativity Thesis must be true as well. For even if it *could* be replaced by an equally effective argument, the *justificatory force* of a thought experiment might still be based on its capacity to make available in a theoretical way those tacit practical commitments which enable us to negotiate the physical world.

2.3.4 Constructivism and the Contrast with Norton and Brown

Before concluding, let me try to articulate more clearly how the position I have been advocating differs from those of James Robert Brown and John Norton, both of whom offer detailed but conflicting discussions of precisely this thought experiment³². Brown's discussions of the case can be found throughout his works³³, but I will focus on his most explicit version, which appears in his book (Brown 1991a), and which is framed partly as a critique of Norton. Norton's own discussion of the case, which is framed partly as a critique of Brown, can be found in his 1996 article (Norton 1996).

Brown contends that the Galileo thought experiment is what he calls a *platonic thought experiment*, that is “a single thought

³¹ For one such (evolutionarily-based) attempt, cf. Shepard 1994; a similar explanation is offered in Sorensen 1992.

³² I thank a non-anonymous referee for the *British Journal for the Philosophy of Science*—John Norton, in fact—for suggesting that I include such a section.

³³ Cf. Brown 1991a, 1991b, 1993b, 1995. In 1991a he speaks of this as “the best (i.e. my favourite)” thought experiment (Brown 1991a, 1).

experiment which destroys an old or existing theory and simultaneously generates a new one” (Brown 1991a, 77, cf. 43-5). According to Brown, thought experiments such as Galileo's give us “*a priori* knowledge of nature” (Brown 1991a, p. 77); through consideration of such thought experiments, Brown maintains, we gain direct quasi-perceptual knowledge of relations between independently existing abstract entities (Brown 1991a, 76). In the case of Galileo's thought experiment, the platonic law of nature to which we gain access is that “all bodies fall at the same rate³⁴.”

Bracketing questions concerning the ontological status of laws of nature, Brown's basic contention is that Galileo's thought experiment brings the heretofore mistaken Aristotelian to see the truth of a certain platonic law of nature, and that it does so neither (a) by introducing the Aristotelian to novel empirical data, nor (b) by showing the law to be a logical truth, nor (c) by making minimal modifications to the Aristotelian's earlier theory, nor (d) by deriving that law from antecedent commitments which the Aristotelian is in a position to maintain³⁵. Of these, (a) and (b) are apparently uncontroversial³⁶, and Brown argues for (c)—whose truth I also accept for reasons enumerated in 2.2.3 above—by contending that the post-thought experimental justification for belief in the Galilean theory is higher than pre-thought experimental justification for belief in its Aristotelian counterpart. The crucial claim, then, is (d) that the Aristotelian cannot be brought to the Galilean conclusion by argument alone.

³⁴ For reasons that Galileo himself addresses, this law is true only of bodies falling in a vacuum; cf. Galileo 1638/1989, 109, which is discussed above.

³⁵ Cf. (1), (2), (3) at Brown 1991a, 77-9.

³⁶ Concerning (a), it is clear that the thought experiment involves no new sensory input, nor is the Aristotelian provided with novel data concerning the behavior of objects in the world. Concerning (b), it is not logically impossible (*simpliciter*) that the speed at which bodies fall should depend on some non-universal feature(s).

It is precisely this, of course, which Norton denies. So before returning to Brown's own positive account of the thought experiment, let us consider Norton's proposed argumentative reconstruction of the case³⁷. Norton suggests that Galileo's reasoning is as follows (Norton 1996, pp. 341-2). The comments in brackets that follow each of Norton's steps are mine.

- (N1) Assumption [for reductio proof]: the speed of fall of bodies in a given medium is proportionate to their weights. [This is equivalent to the negation of (3) in my reconstruction in 2.1; the brackets around "for reductio proof" are my own.]
- (N2) From (N1): If a large stone falls with 8 degrees of speed, a smaller stone half its weight will fall with 4 degrees of speed. [This is an application of the (N1)-law to a particular case. On my understanding of the Elimination Thesis (cf. 2.1.2 above), invocation of this particular case makes Norton's reconstruction—strictly speaking—a thought-experimental argument. But since his reconstruction relies essentially only on the incompatibility of (N1), (N3) and (N5), I will neglect this *ad hominem* point. The reader should be careful, however, not to focus on (N2), (N4) and (N6) in thinking about the forcefulness of Norton's reconstruction.]
- (N3) Assumption: If a slower falling stone is connected to a faster falling stone, the slower will retard the faster and the faster speed the slower. [This is equivalent to (1)—that natural speed is mediative—in my reconstruction in 2.2.1.]

- (N4) From N3: If the two stones of (N2) are connected, their composite will fall slower than 8 degrees of speed. [This is an application of the (N3)-law to the example presented in (N2).]
- (N5) Assumption: The composite of the two weights has greater weight than the larger. [This is equivalent to (2)—that weight is additive—in my reconstruction in 2.1.]
- (N6) From (N1) and (N5): The composite will fall faster than 8 degrees. [This is an application of the (N1)-law to the example presented in (N2) as governed by the (N5)-law.]
- (N7) Conclusions (N4) and (N6) contradict. [As the reconstruction in 2.2.1 demonstrates, a parallel result can be obtained from (N1), (N3) and (N5) alone, without reference to the particular applications that generate (N4) and (N6).]
- (N8) Therefore, we must reject assumption (N1). [Since (N7) is derived from (N1), (N3) and (N5), then one of these three must be rejected. Norton takes (N3)—that natural speed is mediative—and (N5)—that weight is additive—as given, so the only option is to reject (N1). As my discussion in section 2.3 shows, however, there are ways of denying (N3) and (N5) which are—at least *prima facie*—not completely outrageous.]

³⁷ Norton writes: "there can be little question that what Galileo gives us here is simply an argument" (Norton 1996, 341).

Norton then goes on to discuss a final step, which he attributes to Brown.

(N9) Therefore all stones fall alike. [Norton attributes this conclusion to Brown in part on the assumption that Brown accepts that (N8) follows straightforwardly from (N1)-(N7). For reasons that I discuss below, I think this is a misreading of Brown's view. But in any case, as both Norton and Brown recognize, (N9) does not follow from (N8) without additional premises. All that (N8) tells us is that the speed of fall of bodies in a given medium is not proportionate to their weights (that is, the negation of (N1)); it has no bearing on the issue of whether it might depend on some other factor³⁸. Galileo himself is trying to show that "if one were to remove entirely the resistance of the medium, all materials would descend with equal speed" (Galileo 1638/1989, 116), but since precisely what is at issue in the section as a whole is the possibility and nature of motion in a void, this idealization is not permitted as one of the assumptions in the reconstruction.]

Clearly, the move from (N8) to (N9) depends upon an auxiliary assumption:

(N8a) Assumption: The speed of fall of bodies depends only on their weights.

If (N8a) is accepted, then (N9) follows from (N3) and (N5)³⁹; and if (N8a) is not accepted, then (N9) does not. But, Norton notes, Galileo's characters are not in a position to accept (N8a), nor, if we are considering the fall of bodies *in media*, are we. For

³⁸ Cf. Brown: it is not a logical contradiction to suppose that "bodies might fall at different speeds depending on their colours or on their chemical composition" (Brown 1991a, 78).

³⁹ Norton presents this argument as a second *reductio*; cf. Norton 1996, 343.

the speed at which a body falls *in media* is dependent in part on its *shape*: gold leaf falls more slowly than a gold nugget of the same weight (cf. Galileo 1638/1989, 109), just as a closed parachute falls more quickly than an open one. So if we wish to derive (N9) from (N8), we will need to appeal to (N8a).

In short, Norton concludes, "our degree of belief in the final outcome of the thought experiment turns out to depend on our degree of belief in the assumption" (N8a). How so? Well, "the transition from (N8) to (N9) is, at worst, a fallacious inference to a falsehood if (N8a) is denied; or, at best, valid only insofar as it is invoked in special cases in which assumption (N8a) holds, such as the fall of very heavy compact objects in very rare media. This final step looks more like a clumsy fudge or a stumble than a leap into the Platonic world of laws" (Norton 1996, 345). And this outcome, Norton contends, supports the previously-cited thesis that "a good thought experiment is a good argument, a bad thought experiment is a bad argument" (Norton 1991, 131; 1996, 335). For the strength of the thought experimental conclusion correlates directly with the strength of the corresponding argument.

Now, I think the discussion of (N8a) turns out to be a bit of a red herring in the dispute between Brown and Norton, because as far as I can tell, Brown's concession that "the thought experiment is an argument insofar as it destroys the Aristotelian theory" (Norton 1996, 342; cf. Brown 1991a, 78) is not—or at least should not be—a concession by Brown that (N8) follows logically from (N1)-(N7). Like any *reductio*, all that the argument shows is that at least one of the premises needs to be rejected; it does not show us *which* premise. (The point of my discussion in section 2.2.3 was to show precisely this.) So Brown need not concede—as Norton takes him to—that (N8) follows from (N1) [brackets omitted](N7). But, of course, Brown is committed to the view that (N8) is true; indeed, Brown is committed to the view that (N9) is true. So how does Brown think that we know (N8) and (N9), if not by argument from (N1)-(N7) and (N8a)?

Brown's answer, insofar as I understand it, is that thinking about the case forces us (or the Aristotelian) to use our platonic perceptual capacities to see that (N9) is a platonic law of nature (that is, a connection between independently existing abstract entities). But since it seems to me that a perfectly good explanation of what is going on can be offered without appeal to such laws, onto-

logical parsimony suggests that Norton may be partly wrong without Brown being fully right. So let me say what I think is right and wrong in both views as a way of putting my own in context. To give my view a label that contrasts it with Norton's self-proclaimed empiricism and Brown's self-proclaimed platonism, let me call my own position *constructivism* (Kantian associations intended).

I agree with (my reading of) Brown that (N1)-(N7) does not get us (N8); the *reductio* tells us that *something* is wrong with the Aristotelian picture, but it does not tell us *what* is wrong. And I agree with Norton that something is being left out of Brown's story by his neglect of (N8a); indeed, I don't think that the thought experiment in question shows anything more than that natural speed is independent of weight. But what my argument in 2.2.4 aims to show is that the sorts of considerations that allow the conclusion of (N8) from (N1)-(N7) are not available to the Aristotelian as premises before she has followed Galileo's instructions for guided contemplation. It is by focusing on some imaginary particular strapped-body that she comes to realize her non-defeated defeasible commitments to my own (8) (the physical determination of natural speed and weight) and (9) (the lack of physical determination of entification). And therein lies my answer to both Norton and Brown.

Now, if Norton wants to call this sort of guided contemplation *argument*, and Brown wants to call it *platonic intuition*, then we're all in agreement. I prefer to describe it as a reconfiguration of internal conceptual space. By introducing novel categories by which we make sense of the world, this reconfiguration allows us to recognize the significance of certain previously unsystematized beliefs. For the reasons that I discussed in 2.3.2, these beliefs are new; and for the reasons I discussed in 2.3.3, they are justified. Thus the thought experiment brings us to new knowledge of the world, and it does so by means of non-argumentative, non-platonic guided contemplation of a particular scenario.

2.4 Conclusion

In this chapter, I have offered reasons for thinking that a certain view about thought experiments in science is false. The view is that any scientific thought experiment can be reconstructed as a non-

thought-experimental argument without loss of demonstrative force⁴⁰. In Part 2.1, I explained the philosophical motivations for adopting such a view, and distinguished two versions of the position. The first—the Dispensability Thesis—concerns the replaceability of thought experiments; the second—the Derivativity Thesis—concerns their justificatory force. In the remainder of the chapter, I offered reasons for thinking that both of these theses are false. Through a detailed discussion of a thought experiment of Galileo's, I tried to show that the standard argumentative reconstruction of the case fails to capture its justificatory power, and I suggested reasons to think that any other argumentative reconstruction would be likely to fail in similar ways. I then argued that even if one were to provide an argumentative reconstruction that did almost perfectly capture the thought experiment's demonstrative force, this would not show that the *reason* the thought experiment succeeds is because, deep down, it is nothing more than an argument in disguise. I suggested that, to the contrary, the success of the thought experiment may be a result of the way in which it invites the reader's constructive participation, depicts particulars in ways that make manifest practical knowledge, and describes an imaginary scenario wherein relevant features can be separated from those that are inessential to the question at issue.

The final task is to connect this discussion to the broader goals of the study as a whole. One such goal is to establish that guided contemplation of imaginary cases can lead to new knowledge of the world. Showing that this happens in the Galileo case provides me with an example that helps me to establish this conclusion. But there is also a second aspect to the broader project. In Chapters 3 and 4, I will focus more explicitly on the role that exceptional cases should play in the development of theories. And in Chapter 4, where I will contend that something goes methodologically wrong in the far-fetched cases such as those described in the personal identity literature, one of my aims is to pinpoint exactly *where* that is. What I have shown so far is that in some cases, constructive participation in the guided contemplation of imaginary scenarios *can* lead to new knowledge. So although in certain cases this sort of guided contemplation seems to lead us astray, what I have tried to show in these pages is that imagination *as such* is not the problem.

⁴⁰ In 2.3.4 I also briefly offered reasons for rejecting another alternative—that thought experiments offer us special access to platonic laws of nature.

CHAPTER 3

The Ship of Theseus

3.1 CONCEPTUAL THOUGHT EXPERIMENTS

Having examined an example of a factive thought experiment (one where the question being asked was: “what would happen under thus-and-such circumstances?”) we now turn to an example of what I call a conceptual thought experiment. Here, the question being asked is: “how, given that we know what would happen under thus-and-such circumstances, should we describe what happens?” That is, how should the case we are confronted with be accounted for? In particular, this chapter will address instances of how we develop or ascertain or establish criteria for what it is to apply (correctly) the expression “*x* is the same *F* as *y*” when the candidates in question are separated in space or in time¹. We are concerned here with conceptual cases, cases where all of the obser-

¹ In the years since this chapter was first written, there has been an explosion of literature on topics related to the central example, many from perspectives radically different from but complementary to my own. I regret that I have not had the time to modify the central arguments to take account of these discussions (or, indeed, of many valuable discussions already extant at the time of writing.) For a comprehensive listing of relevant works, see the bibliography “Personal Identity and Identity,” which can be found at the end of this volume.

vational data (and some of the non-observational data) have been agreed upon so that, at least in one sense, there is no dispute about what there is².

As before, I will suggest that what gives such thought experiments their justificatory force is the fact that they call upon the reader to perform an *experiment in thought*. She asks herself: “what would I say about thus-and-such a case?”, and then, by observing her own response, she finds out the answer to this question. The answer would be something like: “I would judge that I was confronted with an instance of so-and-so” or “I would feel some pull to say *P* and some pull to say not-*P*” or “I’d be pretty confused and wouldn’t really know what to say.” Here again, performing an experiment-in-thought lets us come to recognize that we have certain beliefs, beliefs which might have remained hidden were they not brought to the fore by the contemplation of this imaginary exceptional case.

In addition, this chapter aims to establish a number of specific claims, the most important of which is this: standard interpretations of the puzzle of the Ship of Theseus have drawn the wrong sorts of general conclusions from the particular case by assuming that the proper way to understand artifact identity is to allow the exceptions to drive the norms, rather than allowing the norms to drive the exceptions. That is, they attempt to resolve the puzzle by making modifications to our pre-reflective identity criteria for objects in general.

Cases where identity is extrinsically determined are cases where whether *x* is identical with *y* depends on something other than the spatio-temporal and causal relations between *x* and *y*. In such cases, our ordinary criteria of identity pick out more than one candidate; they cease working as criteria of identity, and instead become criteria of identity-candidacy. But, I will argue, it is only against a background norm of intrinsically-determined identity

² Note that I am using the term “conceptual” in way that differs from Parfit’s (see Section 3.5.2 below). The sort of case I am considering is one where all observational data are agreed upon, but where there might still be issues about what sorts of underlying facts about the world are producing these phenomena*. As Parfit points out, there is a second sense in which we might speak of a question as being conceptual, namely, in cases where there is full agreement about all the data *and* full agreement about the proper description of the data in terms that all parties agree are the only causally relevant ones; remaining disputes concern nothing more than what all parties would agree are merely issues of *labeling*.

*It might be a further issue of dispute just which data are “observational.” (Thanks to Ted Sider for pointing this out.) I disregard this complication in my discussion above.

that we can make sense of local instances of extrinsically-determined identity³. Most (straightforward) resolutions of the puzzle misidentify the implications of such cases by locating the extrinsic determination in the identity criteria. What I will argue instead is that the best way to account for such cases is to locate the extrinsic determination in the processes by which identity is generally preserved.

3.2 THE STORY

The imaginary scenario that will concern us in this chapter comes in three versions, which I will unimaginatively refer to as “the first version,” “the second version,” and “the third version,” respectively.

First version: There was once a thirty-oared ship belonging to Theseus, which during its seaworthy years underwent gradual repair. Over the years, one by one, each of its original planks was replaced with a new plank of the same size and shape and material, and the old planks were gathered in a barn on the shore, where they rested, piled in a heap. Eventually none of the original planks remained as pieces of the seaworthy vessel, though its appearance was unchanged and its duties unaltered. And although the sophists “of Athens were wont to dispute [whether] after all the planks were changed, [it was] the same numerical ship as it was at the beginning” (Hobbes, *De Corpore II*: 11), Theseus himself faced no difficulties, either practical or metaphysical. He felt no need to (re)christen the craft, no worries about which of the vessels at the dock was his, and no puzzles about whether, strictly speaking, his current ship is the same ship as the ship he commissioned some years back.

Or, to contrast that first version with a *second version*: there was once a thirty-oared ship belonging to Theseus, which during its seaworthy years was disassembled during the winter season. Each fall, when the frosts came, the planks of the ship were pried apart, and stacked in a barn on the shore, where they rested, piled

³ One might take this as (one of the) grounds for endorsing the view that there are no complex material objects (cf., for instance, Unger 1979a, 1979b; van Inwagen 1990; Merricks 1998a, 1998b; Rea 1998; and various of the papers collected in Rea 1997). Since I take it as a given that we do reidentify complex physical objects over time, I instead take it as establishing a (transcendental) precondition for the application of our concepts.

in a heap. Each spring, the pieces were reassembled in a form identical to that of the original ship, so that its appearance remained unchanged and its duties unaltered. And although sophists were wont to dispute its status, arguing that the assembled ship was a different ship than the ship that plied the seas last season, Theseus himself faced no difficulties. He felt no need to (re)christen the craft, no worries about which of the vessels at the dock was his, and no puzzles about whether, strictly speaking, his current ship is the same ship as the ship he commissioned some years back.

Or, to tell the story in a *third version* which is philosophically puzzling: There was once a thirty-oared ship belonging to Theseus, which during its seaworthy years underwent gradual repair. Over the years, one by one, each of its original planks was replaced with a new plank of the same size and shape and material, and the old planks were gathered in a barn on the shore, where they rested, piled in a heap. Eventually none of the original planks remained as pieces of the seaworthy vessel, though its appearance was unchanged and its duties unaltered. One fine afternoon, Theseus collected the planks from the barn, and nailed them together in a form identical to that of the original ship, a form shared by the repaired ship which continued to ply the waters. Suddenly, Theseus faced difficulties, both practical and metaphysical. There were two ships before him: Did one (or both) require rechristening? Which of them was the fine ship of Theseus? Was either of them identical with the ship he had commissioned some years back?

3.3 THE PUZZLE

The perplexities raised by this ancient story are a consequence of the fact that in a world of change, criteria for identity over time permit disruptions in spatio-temporal continuity. In particular,

- (1) objects can survive disassembly and subsequent reconstruction, and
- (2) objects can survive the gradual replacement of component parts over time⁴.

⁴ Cf. Wiggins 1980, 91. Cf. also Hirsch's distinction between what he calls the "soritical rule" (Hirsch 1982, 34–64; cf. Wiggins 1980, 35ff) and the "compositional rule" (Hirsch 1982, 64–71). These are discussed in more detail in Section 3.6.1 below. Cf. *Continued on next page*

Moreover, it seems that objects can straightforwardly survive *both* (2) (replacement of parts) *and* (1) (full disassembly and reconstruction)—so long as the two processes take place in the proper sequence. Suppose, for instance, that an object is first disassembled, then reconstructed, and then has its parts gradually replaced. So long as each of these processes is identity-preserving on its own, no *extra* problem arises from the processes taking place successively. So the problem we face in thinking about the Ship of Theseus is not that the entity has undergone *too many* ordinarily identity-preserving processes to have survived the ordeal; the worrisome feature is not that, with this last process of alteration, the line between sameness and difference has suddenly been crossed.

Rather, the puzzle seems to arise as the result of a particular pattern of events: The two processes are intertwined in a particular way, such that the part-replacement takes place while the object is being disassembled. And here, even though the only sorts of events that have occurred are events whose ordinary result is the (mere) preservation of certain objects, in this particular case, the result of the combination of processes is that at least one new object has been created. This, then, is the puzzle that will concern us regarding the ship of Theseus, namely, *that a process which is ordinarily identity-preserving is in this instance entity-creating*.

Another way to put the problem is this: We need to be able to differentiate cases where the process in question is identity-preserving from cases where it is entity-creating. If we cannot, then given that such a process may be clearly entity-creating (in that it might result in the creation of more than one equally viable candidate), why should we be so sure that it is identity-preserving in the first case? That is, if an intrinsically-specified process could ever be entity-creating, why should we assume that it is ever identity-preserving⁵?

also the distinction between the principle of spatio-temporal continuity of form and the principle of identity of parts (or identity of matter). See, for instance, Smart 1972, 1973; Scatsas 1980; and articles by Burke, Garrett, Lowe, Merricks, Noonan, Rea, and Smart listed in the "Personal Identity and Identity Bibliography". See Bibliography 6.3, page 194.

⁵ Cf. Nozick 1981: There is a principle that "if there could be another thing so that then there would not be identity, then there isn't identity, even if that other thing does not actually exist" (32). (Nozick believes that this principle is false.)

It is precisely this line of reasoning that I wish to block. The Theseus case is exceptional because it describes a circumstance in which an ordinarily identity-preserving process is instead entity-creating. But to conclude from this that a particular sort of process is never identity-preserving, or that identity-criteria in general are somehow suspect, is a mistake. Rather, the correct response to such a case, as I will argue in the final Section 3.7 is to locate the problem in the specification of the process. But in order to get there, I first need to say more about what I mean by identity over time⁶.

3.4 IS THE SHIP OF THESEUS AN EXCEPTIONAL CASE?

3.4.1 Automatic and Specially-Secured Identity

Following Mackie, we might distinguish between “thing-concepts which are such that their ordinary identity-conditions automatically ensure conformity to the logic of identity and those such that special clauses are needed to secure that conformity” (Mackie 1976, 149). The logic of identity, of course, requires reflexivity, symmetry, and transitivity. Everything is identical to itself; to anything that is identical to it; and to anything identical to anything identical to it.

This means, among other things, that if an object at one time is (strictly) identical to an object at a later time, then anything else that the earlier object is identical to is also identical to the later object. Suppose for the moment that we take the two candidates in the third version of the story to be equally good candidates for being identical with the original Ship of Theseus, and, in an effort to be equitable, we say that the original Ship of Theseus is identi-

⁶ Throughout my discussion in this chapter, I am ignoring four-dimensionalist positions (that is, positions which hold that objects are extended in time as well as space). Because I think philosophy should try to make as much sense of our ordinary ways of thinking about the world as possible, I consider such views—which admittedly offer elegant ways of solving many of the perplexities that confront other views—to be a solution of last resort. So this chapter represents an attempt to make sense of our conceptual commitments concerning the status of artifacts as real entities that persist through time. It may be that these commitments are ultimately incoherent and need to be abandoned. But for the purposes of my discussion, I will take them as fixed. (For a defense of four-dimensionalism, see Sider, forthcoming).

cal both to the continuously-repaired ship and to the original-planks ship. Then, given the logic of identity, these two ships must be identical to one another. But, one might continue, since it is evident that they are not, then the original ship cannot be identical to both of them⁷.

Turning next to the first and second versions of the story, we note that in each of these stories the original ship was identical to some ship produced by a process qualitatively identical to one of the processes described in the third version. However, it is fully compatible with (for instance) the first version of the story that, long after the tale has been told, the discarded planks could be reassembled in the form of the original ship. So, Mackie would say, identity in this case is “specially-secured;” we need an explicit no-competitors clause to guarantee that the criterion of identity which the ship meets (continuous replacement of parts over time) actually serves here as a criterion of identity and not as a criterion of candidacy-for-identity.

Mackie argues that such a no-competitors clause is required for any (non-abstract) entity whose continued (diachronic) existence requires more than the persistence of some simple substrate; the ordinary identity-conditions for organisms and artifacts—including ships and trees and human beings—do not offer the guarantee that transitivity will hold⁸. The problem can be illuminated as follows.

Take (Mackie’s reading of) Locke’s general theory of identity through time: that “ x -occurrences at t_1 and t_2 are occurrences of the same x if and only if there is a continuous x -history linking them” (Mackie 1976, 149). That is, two trees at t_1 and t_2 are the same tree only if there is a continuous tree-history linking them, two stereo systems at t_1 and t_2 are the same stereo-system only if there is a continuous stereo-system-history linking them, and so on. The problem that arises is this: tree-histories and stereo-system-histories allow for the possibility that more than one tree (stereo system) at t_2 might be identical with a single tree (stereo system) at t_1 .

For instance, suppose the ordinary identity conditions for trees are that a tree at t_1 is identical with a tree at t_2 if and only if the two bear such-and-such relations. But now suppose that when tree

⁷ Assuming there was only one original ship to begin with.

⁸ See Mackie 1976, 150.

A was a young sapling, it was uprooted, carefully divided into two equal parts, and the two parts—call them *B* and *C*—were transplanted to spots equidistant from the tree's initial rooting. And suppose that both *B* and *C* bear such-and-such relations to *A*. Then by the identity criteria just enumerated, *B* is identical with *A*, and so is *C*. But clearly *B* is not identical with *C*. Since this seems to violate transitivity, the proposed criteria could not suffice for identity⁹. (Obviously, a similar story could be told for the stereo, or, indeed, for any organism or artifact whose identity conditions allow for certain sorts of replacements of parts over time, or for survival under conditions of loss of matter.) For organisms and artifacts, then, it is in general possible that there be two spatially distinct simultaneously existing entities, each meeting the relevant requirements for identity with some earlier entity.

I am concerned in this chapter with a particular sort of case that this makes possible: namely, one where there are two (or more) distinct individuals each of which (would) satisfy the criteria for identity with some earlier individual, but for the existence of the other competitor(s). Mackie's terminology provides a convenient way to describe the difference between cases where such problems might arise, and cases where they cannot. The identity of objects whose identity-conditions are such as to rule out the possibility of identity-candidacy obtaining without strict identity also holding is automatic. The identity of objects whose identity-conditions do allow the sort of multiple-candidacy described above is specially secured.

3.4.2 Organisms, Artifacts, and Exceptional Cases¹⁰

With this distinction in place, let us return to the more general claim I hope to establish: that one of the lessons to be learned from the Theseus case is that it is a mistake to allow our normal ideas

⁹ Alternatively, one might think of this as being a case where symmetry appears to be violated. We might say: *A* is identical with *B* but *B* is not identical with *A*, which would allow transitivity to be maintained. In either case (that is, if symmetry or transitivity is not satisfied) the proposed identity-criterion fails as a strict condition of identity over time. I think locating the problem in transitivity better captures the intuitive sense of what is going on.

¹⁰ In rereading this subsection (1999), it is clear to me that it seriously underestimates the difficulties of the problems it addresses. I retain it here for want of a suitable replacement.

of identity to be driven by exceptional cases. But hasn't Mackie's argument just shown that the Ship of Theseus case is anything but exceptional? Hasn't he established that it fits a pattern that is followed when we enumerate identity conditions for any artifact or organism—namely, that we need to appeal to a no-competitors clause?

I have two responses to offer here; the first concerns organisms, the second—which I will come to in a few pages—concerns artifacts. In the case of organisms, the only sorts of cases where identity-criteria will permit multiple-candidacy are cases where some small proportion of the original mass of the entity (half or less) serves to support the organism's continued existence¹¹. That is, for

¹¹ I am assuming that organisms cannot survive full disassembly and subsequent reconstruction. Van Inwagen concurs: "Note that there is no tendency to identify a 'reassembled' *organism* with the 'original.' If God were to 'reassemble' the atoms that composed me ten years ago, the resulting organism would certainly not be *me*" (van Inwagen 1990, 140; cf. also Penelhum 1957). This is not to say that organisms cannot survive operations in which their organs are removed and replaced, only that they cannot survive being fully disassembled into particle-sized bits, and then reassembled from those bits.

Does this concession raise the possibility of the following Theseus-like case? Over time, my body parts are replaced one by one, and each of the discarded organs is connected up to some sort of machine which enables it to remain functional. After the process of removal is complete, the discarded organs are reassembled in the form of a human being. Is that (a candidate for being) *me*? Here, even though the relevant pieces are pretty large, I am inclined to say that the reassembly process is more like God collecting the atoms that composed me ten years ago than like the surgeon removing my kidney, cleaning it, and replacing it in my otherwise intact body. That is, the process is not identity-preserving. And it is not identity-preserving precisely because it involves full disassembly and subsequent reassembly.

A more difficult version of the case (due to Robert Nozick) would be one in which the machine to which the disassembled parts are connected up is an artificial body, whose parts are gradually replaced over time by the parts that originally belonged to the original human body. Here, the disassembled parts do not need to be reassembled; they are assembled as they are connected up to the machine that preserves them. To the extent that we would be inclined to call the person who ultimately results from this process (a candidate for identity with) the original person, it seems that we are committed to the idea that organisms can survive full disassembly and subsequent reconstruction.

Continued on next page

organisms, it is only under cases of severe mutilation that unsupplemented criteria for being-the-same- x —such as being linked by a continuous x -history—fail to entail strict identity¹². In all other

My inclination is to treat this case as a fission case. Up to a certain point, it resembles the surgeon removing and cleaning the kidney: the original person remains with the original body, while some of his parts have been moved over to a machine which is keeping them functional. At a certain point, however, we come to have a case where a human being has been divided in two: part of (what used to be) him is over here, and part of (what used to be) him is over there. (What we should say about cases like this is something I discuss in Chapter 4.) Eventually, however, these two halves are reunited at the location that was originally occupied by the artificial body. (This sort of situation is generally referred to as “fusion.”) So, I think, this is ultimately a case of the sort that I describe in the main text, where some small proportion of the original mass of the entity serves to support the organism’s continued existence.

¹² But what about cases involving cells which are part of a larger organism, and cases involving entities such as amoebae and worms, all of which regenerate by splitting and for which it seems odd to speak of such cases as “mutilation”?

Replies: (a) Cells can be neglected for the time being on the grounds that they are not strictly speaking *organisms*, but only *parts* of organisms. (And to the extent that they are organisms, the answer I give in (b) applies to them.) (b) Worms and amoebae are organisms, and about them I suppose the right thing to say is this. In part because the way in which such organisms reproduce, identity criteria for such entities are unclear, if we apply ordinary criteria of identity*. So it is not true for amoeba and worms that “it is only under cases of severe mutilation that criteria for being-the-same- x ...fail to entail strict identity,” but the *reason* is not that being cut in half is not severe mutilation for such creatures. Rather, it is because *in general* we do not know what to say about identity over time for such beings, unless we develop special principles meant to apply to entities of those sorts. See also Wiggins 1980, 73 note 20, concerning “wave, volume of fluid, worm, garden, crystal, piece of string, word-token, machine” (italics omitted).

* Cf. Wiggins: “One amoeba becomes two amoebas, but ‘becomes’ receives an analysis making it correspond to ordinary ‘becomes’ as constitutive ‘is’ corresponds to the ordinary ‘is’ of predication and identity. The matter of the original amoeba—the ‘it’—is the fusion, or the matter, of the two new ones taken together. There is matter such that first *a* was constituted of it, and then *b* and *c* were constituted of it” (Wiggins 1980, 72 note 18).

cases, if A is linked by a continuous x -history to B , then B ’s relation to A will conform to the logic of identity. So while identity for organisms may be extrinsically determined (such that we need to look at something besides A and B and the causal connections between them to determine whether A is the same as B) it is only in a highly-constrained subset of the x -history cases that this is so. In particular, it is only in cases where at some point in the organism’s history, it suddenly loses at least half of its matter (and does not thereby go out of existence). So rather than saying, as Mackie does, that identity for organisms is specially-secured, we might say instead that identity for severely mutilated organisms is specially-secured¹³.

But doesn’t this violate conditions of simplicity? Why should we have one set of criteria for applying the concept being-the-same- x -as, and another set of criteria for applying the concept being-the-same-mutilated- x -as¹⁴? My reply: Note first that the difference between these two sets of criteria is trivial; all that the second set has that the first set does not is the clause “so long as there are no equally good competitors.” Indeed, in some sense the proposal is simpler than Mackie’s; he suggests adding this caveat to every enumeration of identity-criteria for organisms; I propose a way of systematically restricting the need for this qualification. Second, Mackie himself acknowledges that any proposal of this sort will seem somewhat *ad hoc*. Special-securement, after all, is a way of connecting a not-so-strict concept to a strict one; it allows us to acknowledge the ways in which entities that we consider to be individuals over time can grow and change in size and shape and composition, without giving up the applicability of strict identity, which demands both symmetry and transitivity. But the lump under the rug needs to go somewhere: in Mackie’s words, “When the concept of one thing of a certain sort is relaxed by allowing growth and so on, it needs to be somewhat arbitrarily restricted

¹³ I discuss such cases (insofar as they concern persons) in the next chapter.

¹⁴ Note that I am not claiming that a mutilated x is not an x . So the criteria for being-the-same-mutilated- x -as will, in some sense, be the same as the criteria for being-the-same- x -as. What I am suggesting is that in the second case, we need an extra proviso to allow us to apply the criteria properly. With this clarification in place, the looseness of expression in the text itself should be less confusing.

elsewhere in order that the resulting ‘thing’ should conform to the strict logic of identity” (Mackie 1976, 171). In this light, my proposed restriction is no more arbitrary than Mackie’s.

Let us turn now to the issue of artifacts. Here, in addition to cases of survival-under-mutilation which parallel the case of organisms and which can be accounted for in the same way, we have the additional possibility that two (or more) competing principles of identity may be at play, such that one of the principles is well met by one candidate, and the other principle by another. Clearly, the Theseus case is a case of this sort: the original-planks ship gains candidacy through the maxim that objects can survive disassembly and subsequent reconstruction, and the continuously-repaired ship gains its candidacy through the maxim that objects can survive the gradual replacement of component parts over time. And here, we cannot employ the strategy employed above; it is indeed normal that artifacts undergo the replacement of parts over time¹⁵, and that they are (occasionally) disassembled and reassembled. So it will not do to carve out a special class of artifacts (as we have for organisms) and say: in those and only those cases do we need to look elsewhere to see whether we have a competitor. For in the case of artifacts, we cannot rule out the possibility of multiple-candidacy merely by observing that the artifact has not undergone fission.

We can, however, rule out the possibility that cases where identity will (as a matter of fact) be specially secured are the rule rather than the exception, by means of a transcendental argument. The argument goes as follows: as the discussion above has shown, we cannot tell by looking at an artifact whether we will have to look elsewhere to determine identity. From this it follows that the cases where we do have to look elsewhere will have to be exceptional. For if they were not, then we would not in general be able to ascer-

¹⁵ By “replacement of parts” I mean something over and above the sort of gradual component replacement that is normal for all physical entities whatsoever. Cf. Shoemaker: “People are sometimes become puzzled by the notion of personal identity on being told that during any seven-year period (or so) all the molecules in a human body are replaced by different ones. Clearly, anyone who is puzzled by the notion of personal identity for this reason should be equally puzzled by the identity of dogs and oak trees...[and] rivers, bicycles...and the like” (Shoemaker 1963, 5). For a discussion of the ways in which this problem troubled the Stoics, see Sedley 1982.

tain artifact identity for artifacts which (like nearly all artifacts) have undergone gradual part-replacement. But this is something that we do all the time¹⁶. So the concept of artifact of which we make use is one that presupposes that specially-secured identity will be the exception rather than the rule.

It may seem, however, that I am letting my epistemology drive my ontology. What does what we can and cannot know have to do with what there is? That we are unable to observe events that take place outside our light-cone does not mean that they do not occur; that an arithmetical sum is too great to calculate in any human lifetime does not mean that there is no fact of the matter. But in the case of artifacts, the issue is more complicated. I will say more about this question in the pages that follow, but let me here note two things that may make the claim more defensible.

First, I take it as uncontroversial that artifacts are interest-dependent; they depend—in ways that non-artifactual entities do not—on the intentions of their maker and the use to which they are put by their employers¹⁷. Second, even leaving us out of the picture, there is something odd about a conception of the world that makes so much of what happens over-here depend on what happens over-there. It is surely true that Cambridge changes sometimes matter in substantial ways¹⁸. If I seek to be the tallest person in the room, then your stepping out the door may allow me to satisfy that description even though I have grown no taller. If you double-fault in tennis, I gain the point even though I have done no

¹⁶ In the sections that follow, I respond to the retort that we cannot, in fact, ascertain artifact identity, either because there are no artifacts (van Inwagen), or because there is no fact of the matter to be ascertained (Parfit). Cf. also Hirsch: “It seems central to the way we think and speak about persistence that we should typically be able to pick out an object and go on to trace its career unambiguously along some space-time path” (Hirsch 1982, 39). And, one might add, that we should be able to do so without keeping our eyes constantly on the object.

¹⁷ Even those who reject the “real” existence of artifacts recognize something to this (Aristotelian) insight. See the discussion of van Inwagen and Rosenberg below (Section 3.5.1.1).

¹⁸ Cf. Geach 1972. A Cambridge change is a change which a thing undergoes as a result of satisfying a description at one time which it fails to satisfy at another, even though the thing has undergone no robust or substantial change. Every object undergoes infinitely many Cambridge changes at any given time, for whenever any

work. And if *A* and *B* are both candidates for identity with *C*, then *A*'s going out of existence may plausibly render *B* identical with *C*, under certain circumstances. But even if such non-local Cambridge changes sometimes matter for identity, it cannot be true that they usually do¹⁹. Even if our view of the world is radically holistic, we nonetheless think that it is sometimes possible to make judgments about individual instances. But if each claim about here and now required a full inventory of the universe (or a reasonably thorough cataloguing of some portion of the universe, or even a clumsy cataloguing of a small part of the universe (we're just not sure which one)), then we would rarely if ever be able to make such judgments; holism would be tantamount to skepticism (or at least to an overly strong form of fallibilism)²⁰. When we speak of identity over time, however, we are not interested in exploring a concept that requires this sort of commitment to a Leibnizian world (where every bit is mirrored in every other). I thus consider any position that would make extrinsic determination the norm to be untenable²¹.

object moves or changes, all other objects undergo Cambridge changes in relation to that object. (And if we allow that the changes in question can themselves be Cambridge changes, then each change ramifies infinitely. Each Cambridge change in each object brings about a corresponding Cambridge change in every other object, each of which in turn brings about a corresponding Cambridge change in every other object, and so on.)

¹⁹ Derek Parfit points out that on his view, *all* questions of identity are Cambridge questions. If we define non-Cambridge properties as properties which can play causal roles (in non-intentional contexts), then identity seems clearly to be a Cambridge property. Indeed, it was central to our discussion of the Galileo case (Chapter 2) that questions of identity generally go unanswered by the world. I do not mean to deny this in my discussion here. The aspect of Cambridge changes with which I am concerned is their *non-locality*.

²⁰ This is not to deny that there is an important strand in philosophy (and certainly in religion as well) which stresses our insurmountable ignorance as finite beings.

²¹ (Note added in 1999.) Although I am still committed to something like the conclusion presented here, it is clear that the arguments in this Section (3.4.2) are not sufficient to establish it.

The purpose of this sub-section has been to clarify what I mean when I say that the case of the Ship of Theseus is an exceptional case. With this in place, let us turn to a number of influential interpretations of the case's import. These interpretations can be fruitfully divided into two categories. Interpretations of the first type attempt to dissolve the problem by showing that although there seemed to be something deep at issue here, the puzzle actually rests on false assumptions about the relation of our concepts to the world. Interpretations of the second type attempt to solve the problem by coming up with general principles which can cover the case. After presenting the interpretations, I will try to show why they miss what I think is really at issue, and why this is a consequence of mis-generalization from an exceptional case.

3.5 ATTEMPTS TO DISSOLVE THE PROBLEM

3.5.1 Van Inwagen

In *Material Beings*, Peter van Inwagen advances a thesis which he calls “the Denial”: that “there are no tables or chairs or any other visible artifacts except living organisms” (van Inwagen 1990, 1)²². And “if there are no artifacts, then there are no philosophical problems about artifacts.” In particular, there are no problems with “identity and persistence through mereological change” (van Inwagen 1990, 128). So the puzzle of the Ship of Theseus is easily solved; “there are no ships, and hence there are no puzzles about the identities of ships” (van Inwagen 1990, 128). “All that happens in the story is that planks are rearranged, shuffled, brought into contact, separated, and stacked²³. But at no time do two or

²² Or again: “My position *vis-à-vis* tables and other inanimate objects is simply that there *are* none. Tables are not defective objects or second-class citizens of the world; they are just not there at all...There are certain properties that a thing would have to have to be properly called a ‘table’ on anyone’s understanding of the word, and nothing has all of these properties” (van Inwagen 1990, 99-100). Or yet again: “There are...no tables and chairs, and there are no other artifacts. Artisans do not create; not, at least, in the sense of causing things to exist. They rearrange objects in space and cause bonding relations to begin to hold or cease to hold...between objects” (van Inwagen 1990, 127).

²³ In telling the story, van Inwagen treats planks as “honorary simples.” He suggests that one might speak more precisely of “simples arranged plankwise” (van Inwagen 1990, 128).

more of these planks compose anything, and no plank is ever a proper part of anything...[T]he story end[s]as it beg[ins]: with no ships at all" (van Inwagen 1990, 129).

Despite his hard-nosed attitude about what there really is, in the ordinary conduct of things, van Inwagen is perfectly willing to speak of artifacts²⁴. What he suggests is that there is a certain mismatch of languages when, on the one hand, we refer to such entities as artifacts while, on the other hand, we insist on holding our talk to standards of precision appropriately reserved for philosophical conversation. He writes: "I am just as willing as you are to use sentences containing 'house' or 'ship' in the ordinary business of life. But if you begin to insist that the things we are talking about strictly and in every respect conform to such general logical principles as the Law of Excluded Middle, then I shall insist that we have departed from the ordinary business of life, and I shall consequently insist that we adopt a language capable of bearing [such] weight...a language that refers to nothing besides simples and living organisms and abstract objects" (van Inwagen 1990, 131)²⁵.

But so far, this gets us nowhere in terms of the original puzzle. Presumably, van Inwagen would allow that common parlance permits us to speak of the continuously-repaired ship in the one-sided case (where the original planks are destroyed) as being the same ship as the original Ship of Theseus. And, presumably, common

²⁴ Van Inwagen suggests that talk of such entities be understood on analogy with talk of the sun's moving. Just as it is compatible with a Copernican world-view to say : "It was cooler in the garden after the sun had moved behind the elms," so too is it compatible with the Denial to say: "Some of my chairs are being re-upholstered in eggplant-colored velveteen." (For further discussion of this analogy, see van Inwagen 1990, 1-2 and 98-107).

²⁵ Again, he offers an analogy: "If you get sticky about strict adherence to the Principle of Noncontradiction-if, that is, you insist that even the surface structure of my sentences never be of the form 'p and not p'-then I shall stop saying 'It is and it isn't' in response to 'Is it raining' and instead talk of mists. If you insist that every piece of language that has the syntactical form of a singular referring expression denote an object and that every predicate expression concatenated with such a referring expression express a possible property, I shall stop saying 'The average father has 1.3 children' and say that the number of children divided by the number of fathers is 1.3" (van Inwagen 1990, 131). See also van Inwagen 1990, 98-107.

parlance allows us to speak of the disassembled and reconstructed ship in the one-sided case (where no repairs are made) as being the same ship as the original Ship of Theseus. So the question is: what does common parlance permit us to say in the case described by the third version? That is, simply pointing out that there are no ships does not help us with what I have identified above as the core of the puzzle, namely that the Theseus case presents us with an instance where a process that is ordinarily identity-preserving (where by this, we mean: identity-preserving in the common parlance sense) is in this instance entity-creating (where again, what we mean by this is: entity-creating in the common parlance sense). Pointing out that, strictly speaking, there are no ships does nothing to help us in making headway with this problem²⁶.

But where van Inwagen's suggestion is helpful is in freeing us from applying inappropriately strict standards to the answer that we give in this case. Perhaps when we speak of identity in the common-parlance sense, "identity" can be intransitive, so that the original Ship of Theseus can be said to be identical with each of the two final candidates, but they are not identical with one another. Or perhaps when we speak of identity in the common-parlance sense, identity can be partial, so that the original Ship of Theseus can be said to be somewhat identical with each of its successors, but not fully so. Or perhaps when we speak of identity in the common-parlance sense, identity can depend on interest, so that the reconstructed ship can be said to be identical to the original Ship of Theseus if one is interested in the question for antiquarian reasons, and identical to the repaired ship if one is interested in the question for marinarial reasons. Or perhaps when we speak of identity in the common-parlance sense, identity over time is subject to extrinsic constraints, such that what happens to C can play a role in determining whether A and B are identical, even going so far as to make A identical to B until the moment that C comes into existence.

²⁶ The problem I am pointing out connects to a general difficulty faced by such ideal-language approaches. One lesson of positivism's legacy is that replacing natural language with some highly-constrained artificial language is problematic for at least two reasons: (a) that natural language does not seem to be the jumbled mess that radical critics take it to be, and (b) that if it were, it is unclear how we could come up with the conceptual resources necessary to replace it with something more precise.

3.5.1.1 Identity Under a Sortal

One popular way of dealing with the puzzle, which makes use of the insight we have been considering in the last paragraph, is to distinguish the two candidates on the basis of interest. Such a solution can be neutral with regard to the question of whether ships and so on really exist. For instance, van Inwagen writes: “Statements that are apparently about the continued persistence of artifacts make covert reference to the dispositions of intelligent beings to maintain certain arrangements of matter” (van Inwagen 1990, 134), while Rosenberg, in direct challenge to van Inwagen writes: “correct answers to questions regarding the identities of artifacts turn on facts regarding the intentional activities of persons” (Rosenberg 1993, 708). So regardless of our ontological commitments, we might say that the original ship is the same as the continuously-repaired ship if we are interested in ways that the arrangement of matter is maintained in order to promote seaworthiness, and that it is the same as the reconstructed ship if we are interested in the ways that the arrangement of matter is maintained in order to promote antiquarian precision. (And we might expect, moreover, that if we fail to specify which one concerns us when we use the term “ship”, it is likely to be the former rather than the latter²⁷.)

In many ways, this answer is appealing; it seems to capture the reasons why we are inclined to take each of the ships in the third version of the story to be viable candidates for identity with the original ship. And it seems to show that the conflict results from a conflict between standards, not a conflict within a single standard. Consider the story by David Wiggins wherein a monument is to be erected in honor of Theseus, on the top of which will be placed his ship. Wiggins suggests that:

²⁷ Mackie suggests that this would be Locke’s solution. He writes: “Locke’s theory would lead us to say that the seaman’s ship [the ship which has resulted from the constant repairs] is the same *ship* as the original one, since there is a continuous ship-history linking them, whereas the antiquarian’s ship is only the same collection of ship components as the original, since what links them is a continuous history of what has for most of the time been a dispersed collection of ship components, and such a dispersed collection is not a ship—particularly when throughout much of this time a large part of this collection has formed part of something else that was undeniably a sea-going ship*” (Mackie 1976, 143).

*Note, however, that adding the qualification “particularly when...” is not as innocent as it initially seems.

Surely some people would say that the ship put together from discarded planks was the right one to raise up there. And a dispute might break out about this matter between priests who favoured the working ship and antiquarians who favoured the reconstruction. The difficulty is then a certain incomparability to their positions. It may seem that one party would be looking for an archaeological relic and the other for a functionally persistent continuant; and that the dispute was to be traced to a disagreement about what it is for something to be a sacred ship. The antiquarian who favours the reconstructed ship has a different interest, it might be said, from the priest who favours the continuously repaired continuant. Both are stuck with the qualification ship, but they have different interests. (Wiggins 1980, 93-94)²⁸.

So it seems that we have solved the puzzle: the repaired ship is the same sailing-ship as the original, and the reassembled ship is the same museum-ship as the original, and that’s that²⁹. The conflict was only apparent; metaphysics has been saved by a couple of carefully-placed hyphens.

But on reflection, it is not so clear that we really have a solution. For suppose that we return to the first story, where the parts of the ship are replaced over time, and the original planks are somehow destroyed. And now suppose that a monument is to be built in honor of Theseus. Surely the antiquarians would not say: “What a hopeless endeavor! There will be no appropriate ship to put atop the pillar, for the brave ship of Theseus has been destroyed piecewise over the ages³⁰.” And the reason that they would not say this is not because they have, as a second meaning for ‘ship’, the meaning that the priests have. The reason is that even if one’s interests are antiquarian, artifacts can survive the gradual replacement of parts over time.

²⁸ Wiggins does not think that this retelling provides a full solution to the puzzle. He continues: “we must not confuse the fact that it is in some sense a psychological matter whether *we* adopt the priests’ or the archaeologists’ view with it being a merely psychological matter which one is Theseus’s ship—or an arbitrary matter” (Wiggins 1980, 94).

²⁹ And the original ship? It was, well, um, both. (See next paragraph.)

³⁰ Note that were they committed to the view that the original *matter* had to be the same, then they would be unable to countenance the persistence of *any object whatever*, since all objects undergo a natural process of replacement of parts over time (cf. note 15). Presumably, however, one might insist that this process be a *natural* one.

A similar argument applies in the other direction. Suppose that we return to the second story, where the planks of the ship are disassembled during the winter months. Suppose that some of them are even used as see-saws and balance beams at a large indoor playground. And suppose again that a monument is to be built in honor of Theseus. Surely the priests would not say: "But wait! What do you think you are doing? There's nothing to put atop the monument. That old ship went to pieces many winters ago." And, again, the reason they would not say this is not that they have, as a second meaning for 'ship', the meaning that the antiquarians have. The reason is that even if one's interests are functional, artifacts can survive disassembly and reassembly³¹.

3.5.1.2 Summary

Let me summarize where we stand so far. Van Inwagen has argued that, strictly speaking, there are no artifacts. When we speak of artifacts persisting, we are really speaking of the dispositions of intelligent beings to maintain certain arrangements of matter. I suggested that we can accept the core of this insight, that the persistence of artifacts is essentially tied to the interests of intelligent beings, while remaining neutral with respect to van Inwagen's ontology. One consequence of endorsing this aspect of van Inwagen's view is that there is a certain weakening of strictures concerning what can be said about objects. Among other things, it

³¹ Nor is the problem that "ship" is somehow ambiguous, in the way that "passenger" is. We might imagine that the ship set sail to Delos in the morning with *A*, *B*, *C* and *D* aboard, and that it returned in the afternoon with *A*, *C*, *E*, *F* and *G*. And we might ask: "how many passengers were there on the Ship of Theseus today?" The answer to this question is *nine* under one construal of *passenger* (*A*, *B*, *C*, and *D* on the way out, and *A*, *C*, *E*, *F*, and *G* on the way back), and *seven* under another (*A*, *B*, *C*, *D*, *E*, *F*, and *G*). (Thanks to Zoltán Gendler Szabó for discussion here.)

Could this analogy help us with our two apparent readings of ship? I think not. The problem with the Ship of Theseus is not that, on one reading, ships can be counted on the basis of antiquarian considerations, and on another, they can be counted on the basis of functional considerations. Nor is the problem that we do not have a sufficiently precise idea of how to count ships*. The problem is that, within a certain range, even if we fully specify what we are interested in when we speak of ships, both criteria (disassembly-reassembly and the replacement of parts over time) seem to be perfectly good criteria for identity-preservation.

*Cf. Wiggins 1980, 73: "there is no universally applicable definite way of counting crowns. The Pope's crown is made of crowns, There is no definite answer, when the Pope is wearing his crown, to the question 'how many crowns does he have on his head?'"

may be true that there can be two distinct artifacts in the same place at the same time, without that implying that single spaces can be multiply occupied by interest-independent entities (such as rocks)³². This opens up the possibility of thinking of the puzzle of the Ship of Theseus as being a puzzle that concerns an ambiguity in the sortal ship, such that on one reading, a ship is something that serves the needs of mariners, and on the other, a ship is something that serves the needs of curators. But I suggested that this is not a satisfactory solution to the puzzle, since even if one's interests are practical, it seems that artifacts can survive disassembly and subsequent reassembly, and even if one's interests are antiquarian, it seems that artifacts can survive the gradual replacement of parts over time.

3.5.2 Parfit

On the surface, it seems that nothing could differ more from van Inwagen's position than a position which grants existence to just about everything. Van Inwagen is extremely ontologically parsimonious, and the position I am about to describe represents an extreme in ontological profligacy. Nonetheless, I will contend, as far as the puzzle of Theseus is concerned, the two positions offer precisely the same resources for providing a satisfactory analysis of the puzzle (and neither is adequate).

Parfit contends that "if the existence of *Xs* just consists in the existence of *Ys*, *Xs* are not, in relation to *Ys*, separately existing entities" (Parfit 1994, 9). So, for instance, suppose that we follow Parfit in saying that "the existence of a library just consists in the existence of a building and a set of books, and in the activities of various people" (Parfit 1994, 3). If we let the *Xs* be libraries and *Ys* be buildings, books, and activities of various people, then we have an instance where *Xs* (that is, libraries) are not separately existing entities in relation to *Ys*. In a parallel fashion, we might say that the existence of ships (*Xs*) just consists in the existence of planks (arranged in a certain fashion and used for a certain purpose) (*Ys*); so ships, in relation to planks, are not separately exist-

³² (Note added 1999.) This argument goes by far too quickly. Without some sort of eliminativist ontology (and/or some extremely sharp line between natural and non-natural objects), you can't just have artifacts for free. (Thanks to Ted Sider for pressing me on this.) Despite this rather glaring omission, the negative argument still goes through.

ing entities. For the sake of convenience, let us speak of the Ys as the basis for the Xs.

Parfit also discusses a category of entities which he calls *notional*. Notional entities are entities which are *mind-dependent*, entities which depend for their existence as *entities* on our thinking of them as such. So, for instance, the constellation *Ursa Major* is mind-dependent in this sense. Even though the stars which serve as its basis would exist regardless of our activities, and even though they would be arranged in the pattern which they are now arranged, the existence of the *constellation* depends upon our thinking of those stars as forming that pattern. What it is for something to be a constellation is for there to be some arrangement of stars in the heavens to which we assign a label; so the very idea of constellation has built into it some sort of observer-dependence, even though the configuration of stars which makes up the constellation does not³³.

Parfit contrasts this sort of mind-dependence with the mind-dependence of entities such as fictional characters. Fictional characters, he suggests, are brought into existence by our conceptual activities. It is not a matter of our giving a label to some substrate that exists regardless of our interaction with it; when we invent fictional characters, we create the Xs from scratch. So, he concludes, most ‘notional’ (that is, mind-dependent) beings are not “brought into existence by our conceptual activities. They would exist whatever concepts we employed. But their existence might be claimed to be, in another sense, conceptual” (Parfit 1994, 10). Ships and forests and routes are notional entities in this sense³⁴. Parfit is perfectly willing to grant existence to any notional entity whose basis exists. As far as he is concerned, there really are ships and forests and routes.

³³ Of course, the stars themselves may be (in this sense) observer-dependent; it may be that we see them as discrete only as a result of our concepts. But I take it as uncontroversial that stars are (in some sense) *less* observer-dependent than constellations.

³⁴ Cf. Parfit: “Suppose I already know that several trees are growing together on some hill. I then learn that because that is true, there is a copse on this hill. That would not be new factual information. I would merely have learnt that such a group of trees can be called a ‘copse’. That is a conceptual fact. And it provides a sense in which, in relation to the trees, the copse is a ‘notional being’. Though...

Continued on next page

He goes on to point out that in certain cases, even though we are fully cognizant of the underlying facts, we might still ask questions about notional entities. Suppose, for instance, that I gaze out the window and ask: “Is there a copse on that hill?” I might be asking one of two questions. I might be wondering whether the green circle which I see on the hillside is a cluster of trees or a celebratory tent. Or I might be wondering whether those five or six trees which I see over there really constitute a copse. If my worry is the first, I am asking a *factual* question; if my worry is the second, I am asking a *conceptual* question³⁵.

Now, what Parfit would presumably say about the puzzle of the Ship of Theseus is that the question “which of these ships is the Ship of Theseus?” is a conceptual question. We already know everything there is to know about the underlying facts: we know what happened to each of the original planks, and we know how each of the two competitors came into being. Our worry is not about which of two possibilities will be realized. Our worry is about how we should describe a state of affairs given that we already know everything there is to know about the basis.

But at this point, we find ourselves exactly where we were when we were told by van Inwagen that the worry that confronts us in the case of the Ship of Theseus is a worry about common parlance. Our rules for answering conceptual questions about ships permit us to speak of the continuously-repaired ship in the one-sided case

copse do not exist only because of the way we talk, when we learn that copses exist, we may only be learning a fact about how we talk” (Parfit 1994, 10), and Chisholm: “[T]here is no dispute about any observational data. You have agreed about what it is that is called ‘Elm Street’, about what it is that is called ‘Route 42’, about the number of lanes in the various places, and about what parts are composed of what. Your dispute, then, has to do with criteria for *constituting the same road*.” (Chisholm 1971, 4-6). There is, of course, a second way in which ships (and most roads) are interest-dependent and forests are not: ships are constructed artifacts. (That there are planks assembled in this particular ship-wise pattern is the result of some intentional agent having arranged them that way.) I am bracketing this issue for the moment, along with the vexed question of what might be called *found artifacts*, that is, previously existing entities that are put to a particular use by some intentional agent.

³⁵ “Even when a question is about reality, the answer depends not only on reality but also on our concepts. If we already know about the former, we must be asking about the latter” (Parfit 1994, 11).

(where the original planks are destroyed) as being the same ship as the original Ship of Theseus. And, presumably, our rules for answering conceptual questions about ships allows us to speak of the disassembled and reconstructed ship in the one-sided case (where no repairs are made) as being the same ship as the original Ship of Theseus. So the question is: what do our rules for answering conceptual questions about ships allow us to say in the case described by the third version? Simply pointing out that questions about ships are questions about the application of our concepts does not help us with what I have identified above as the core of the puzzle, namely that the Theseus case presents us with an instance where a process that is ordinarily identity-preserving (in these sense that our concepts would lead us to say that an object has persisted) is in this instance entity-creating (in the sense that our concepts would lead us to say that an object has been brought into existence). And from here, the argument continues as in the last section. Even if the question at hand is only a question about our words, it remains puzzling how it is that a practice that is ordinarily such that it allows us to track identity might sometimes allow us only to track identity-candidacy.

3.6 ATTEMPTS TO SOLVE THE PROBLEM

3.6.1 A Traditional Solution: Hirsch

Traditional characterizations of the puzzle of the Ship of Theseus are characterizations that suggest that the puzzle arises when two criteria of identity are pitted against one another. In particular, the standard presentation suggests that the criterion of continuity of form is pitted against the criterion of continuity of matter, such that our intuitions pull in both directions. Traditional solutions to the puzzle of the Ship of Theseus are solutions which come down in favor of one or another of these two principles, concluding that one or the other (or neither) of the two resulting ships is the Ship of Theseus, on the grounds that one or the other (or neither) of the competing principles takes primacy in this case.

So, for instance, Eli Hirsch's interim solution to the puzzle falls into this category³⁶. Hirsch suggests that the puzzle arises when "considerations of compositional similarity and continuity yield

³⁶ I discuss his considered solution in Section 3.7 below.

"incompatible identity judgments" (Hirsch 1982, 67) and he comes down, albeit somewhat tentatively, in favor of the constantly-repaired ship.

In order to understand the remainder of Hirsch's discussion, we will need two pieces of terminology. Hirsch's *Sortal Rule* (sufficient condition version) says:

The Sortal Rule: A sufficient condition for the succession S of object-stages to correspond to stages in the career of a single persisting object is that:

- (1) S is spatiotemporally continuous³⁷; and
- (2) S is qualitatively continuous; and
- (3) There is a sortal term F such that S is a succession of F-stages(Hirsch 1982, 36)

And his *Compositional Criterion* (second formulation) says:

The Compositional Criterion: Where x is an object that exists at time t_1 and y is an object that exists at a later time t_2 , a sufficient condition for x to be identical with y is that the same sortal is predicatively true of both x at t_1 and y at t_2 , and some set of objects comprises a major portion of both x at t_1 and y at t_2 , and this set of objects is similarly arranged in both x at t_1 and y at t_2 (Hirsch 1982, 65).

Informally, the Sortal Rule says that an object can persist through changes so long as the spatiotemporal continuity is not disrupted, and so long as the changes it undergoes are characteristic of entities of a certain sort, of which this object is one. The Compositional Criterion says that an object can persist through disassembly and reconstruction, so long as the earlier and later candidates are both composed of roughly the same matter arranged in roughly the same way. So the Sortal Rule can be seen as a more precise expression of what we have been calling the second principle, and the Compositional Criterion can be seen as a more precise expression of what we have been calling the first principle.

³⁷ A characterization of this notion can be found at Hirsch 1982, 15-22.

After presenting the story of the Ship of Theseus, Hirsch writes: “My own somewhat ambivalent inclination when reflecting upon this case is to judge that [the original ship] is identical with the [constantly-repaired ship] and not with the “[reconstructed]” ship (Hirsch 1982, 69). He continues: “If this intuition is generally shared there would be two related ways to explain it. It will be noted that the judgment that [the original ship] is identical with [the constantly-repaired] follows from the sortal rule as well as from the compositional criterion, whereas the judgment that [the original ship] is identical with [the reconstructed ship] follows only from the compositional criterion... We might say that... where the sortal rule conflicts with the compositional criterion the former rule, which we know to be primary, takes precedence. Or, forgetting about primacy, we might simply say that the sortal rule in conjunction with the compositional criterion outweighs the latter standing alone” (Hirsch 1982, 69). So Hirsch’s tentative solution to the puzzle is to grant identity with the original ship to the constantly-repaired ship, and to do so on the grounds that the criteria by which the latter can be said to be identical with the former are criteria which more precisely characterize the grounds on which identity over time can be said to obtain.

How successful is this solution, given our initial question? The initial question, you will recall, is what should be concluded from the fact that the Ship of Theseus story seems to present us with the story where a process that is usually identity-preserving is in this case entity-creating? What the traditional answer says is this: “Yes, such a process is usually identity-preserving. But it is not identity-preserving here. And the reason it is not identity-preserving is that there is another, better process which gives us another, better candidate.” So the traditional answer is tantamount to saying that in this case, the compositional criterion is not really a criterion of identity; it is merely a criterion of identity-candidacy, a second-best way of being identical over time.

Such a solution is unsatisfying because it is either too general or too *ad hoc*. It is too general if what the solution says is that whenever an object persists by means of the compositional criterion, all we can conclude is that we have an persistence-candidate, for this is surely false. And it is also too general if the solution says that whenever the two rules conflict, we should choose the sortal rule, for then the position is subject to the sorts of objections that I make against the closest continuer theory (see Section 3.6.2.2

below). On the other hand, if the solution says that there is something about this case of conflict which gives the sortal principle criterion primacy over the compositional, then it is too *ad hoc*, for there is nothing explanatory about such a diagnosis. And such a criticism will apply to any solution which gives priority to one or the other principle without identifying some dynamic relation between them; either it will falsely denigrate the status of the principle which fails to receive priority, or it will fail to explain why this case is a case where the principles in question conflict.

But as before, there is an aspect of this proposed solution that gets at something deep about the puzzle of the Ship of Theseus. It is true that there is something odd going on when the two principles come in conflict with each other. And it seems right to say that but for the candidate presented by the second principle, the process identified by the first principle would be identity-preserving. So there is something in the interaction between them that is causing the conflict. To this extent, the traditional solution is after something deep. But it does not yet offer the full answer.

3.6.2 A Meta-Solution: Nozick

In *Philosophical Explanations*, Nozick contends that “to be something later is to be that thing’s closest continuer;” to be something later is to be the entity that most closely satisfies the set of criteria by which the original entity may be generally said to continue over time. So to be E at t_2 is to be the entity which best matches those characteristics which, at t_1 , would be expected to hold of an entity that followed the natural trajectory of the (ideal) E from t_1 to t_2 , and that matches those characteristics for reasons that depend causally on E³⁸.

So, for instance, let us suppose that the Vienna Circle in the mid-1930s had twenty members, who met weekly in Vienna to discuss philosophy³⁹. According to the natural trajectory of the ideal Circle, we might expect that in 1945 it would have the same twenty members, that these members would live in Vienna, and that the

³⁸ “To say that something is a continuer of x is not merely to say its properties are qualitatively the same as x’s, or resemble them. Rather it is to say they grow out of x’s properties, are causally produced by them, are to be explained by x’s earlier having had its properties, and so forth” (Nozick 1980, 35).

³⁹ The example is from Nozick (1981), 32-33; I have added certain details.

group would have met weekly to discuss philosophy during the intervening decade. Instead, however, what we find in 1945 are the following candidates for identity with the original Circle: (a) a trio of members who spent the War in Istanbul, meeting weekly; (b) two members who spent the War hiding in Vienna, meeting only twice during the period from 1939 to 1945; and (c) nine members who spent the War in the United States, meeting twice each month.

Each of these candidates meets some of the initial criteria; in particular, (a) met weekly during the decade from 1935 to 1945, and (b) has members who live in Vienna. Nonetheless, it seems that (c) has the best claim to the title "Vienna Circle." The group, which was composed of a good proportion of the original members, met regularly to discuss philosophy; in this light, the Vienna criterion seems to hold less weight. However, as Nozick points out, it also seems right to say that, were (c) not to exist, (a) would be the Vienna Circle. "If no other group exists, the Istanbul group is the closest continuer; but if the group in the United States exists, it is the closest continuer" (Nozick 1980, 33). Since what it is to be the Vienna Circle at t_2 is to be the closest continuer of the Vienna-Circle-at- t_1 , then the Istanbul group would, but for the existence of its New World competitor, be the Vienna Circle.

Let us look first at how this theory applies to the case of the Ship of Theseus, where it seems initially to account for our intuitions extraordinarily well, and then examine some of its counterintuitive implications. Since the closest continuer theory is a schema, it does not tell us which of the two competitors in the third version is the Ship of Theseus. What it does instead is to provide us with a framework for making sense of the puzzle⁴⁰. The theory explains why we should expect to encounter cases with the structure of the Theseus case, that is, cases where information about the existence of some object C leads us to deny the status of continuer-of-A to some wholly other object B. Nozick writes:

In the case of the ships, there are two relevant properties: spatiotemporal continuity with continuity of parts, and being composed of the very same parts (in the same configuration).

If these have equal weight, there is a tie in closeness of con-

⁴⁰ "The closest continuer theory does not, by itself, answer the question...[but] it helps to sort out and structure the issues" (Nozick 1981, 33).

tinuation. Neither, then, is the closest continuer, so neither is the original ship. However, even when the two properties receive equal weight, if there had actually been one ship existing without the other, then it, as the closest continuer, would be the original ship...[So] the closest continuer schema does fit and explain our response to this puzzle. When we hear the first story...we are not puzzled or led to deny that it really is the same ship. Only when we learn of the reconstituted ship are we thrown into puzzlement, not only about its status, but about the earlier product of gradual rebuilding. *It is only when we learn of another candidate for closest (or equally close) continuer that we come to doubt that the gradually altered ship is the same ship as the original one* (Nozick 1981, 33-34, italics added).

There is something extremely appealing in this diagnosis. It does seem that it is only when we learn of the existence of the second ship that we come to doubt the status of the first. Moreover, such a theory helps refine the sortal-relative solution presented above in the context of Wiggins's story of the priests and the antiquarians⁴¹. As I pointed out above, the problem with the diagnosis that suggests there are two different concepts of ship, one antiquarian and one practical, is that the diagnosis cannot explain why subscribers to either concept would be inclined to accept the validity of the alternative. But the closest continuer theory can, and it can do so while preserving the insight that part of what is going on is that there are various interests at play. What the closest continuer theory can say about the priests and the antiquarians is that they have different rankings of the criteria for continuity. Each side recognizes that the other's principle has some force, but the one side thinks reassembly trumps replacement, while the other thinks the opposite.

But despite its considerable appeal in this case, the closest continuer theory has at least three extraordinarily counterintuitive implications. And these counterintuitive implications are, I think, sufficient to override its otherwise alluring aspects.

⁴¹ The reader will recall that the priests are those who believe that the Ship of Theseus is the repaired ship, on the grounds that they are concerned with sailing-ships, and that the antiquarians are those who believe that the Ship of Theseus is the reassembled ship, on the grounds that they are concerned with museum pieces.

First, on the closest continuer theory, identity over time is not transitive: *A* at t_1 may be identical with *B* at t_2 and with *C* at t_3 , but *B* and *C* may not be identical. Suppose, for instance, that *A* is our original entity. And suppose that, at t_2 , *B* is its closest continuer. So at t_2 , *B* is *A*. Suppose, however, that at t_3 , *C* is *A*'s closest continuer, but that the closest continuer of *B* is *B'*. Then at t_3 ; *C* is *A*, and *B'* is *B*. But *B'* is not *A*⁴².

An example may help to make this clearer:

- t_1 : Let *A* at t_1 = {Albert, Bernadette, Camille, Dorothy, Ella, and Fred}.
- t_2 : Let *B* at t_2 = {Albert, Bernadette, Camille}
Let *C* at t_2 = {Dorothy, Ella}
Let *D* at t_2 = {Fred}.
- t_3 : Let *B'* at t_3 = {Albert, Bernadette, George}
Let *C'* at t_2 = {Camille, Dorothy, Ella}
Let *D'* at t_3 = {Fred}

At t_2 , *B* is *A*. It is a good enough continuer of *A*. And it is the closest continuer of *A*, since it has three of *A*'s original members, while *C* and *D* have only two and one respectively. At t_3 , however, *C'* is *A*. It is a good enough continuer of *A*. And it is the closest continuer of *A*, since like *B*, it has three of *A*'s original members. But *C'* is *not* the closest continuer of *B*; it shares only one member with *B* (Camille), while *B'* shares two (Albert and Bernadette). So we have the following relations

- At t_2 : *B* = *A*
- At t_3 : *B'* = *B*
- At t_3 : *C'* = *A*
- At t_3 : *C'* ≠ *B*

That is, at t_3 , *C'* is not identical to *B*, but it is identical to *A*, which is (at t_2) identical to *B*⁴³.

The second counterintuitive implication of the closest continuer theory is that objects may move instantaneously from one

⁴² For another example of this sort (which differs in that *A* has no continuer at t_3), see Nozick 1981, 659, note 10.

⁴³ Nozick, of course, is well aware of this implication. He writes: "this...relation need not be transitive," and goes on to present four ways that "a view of identity [might] cope with these nontransitivities of...closest continuer" (Nozick 1981, 42-43).

location to another, without being anywhere in between: again, *A* at t_1 may be identical with *B* at t_2 and with *C* at t_3 , where *B* and *C* are spatially distant from one another. Continuing with the previous example, let us label *A* "The Club", and let us suppose that, at t_3 , The Club is in Idaho, where we find Camille, Dorothy and Ella. At t_4 , the three die tragically in a boating accident, and *B'* (that is, the group: Albert, Bernadette and George) becomes The Club. But *B'*, let us suppose, is in Indiana. So at the moment of the boating accident, The Club moves from Idaho to Indiana, instantly and without being anywhere in between⁴⁴.

Third, according to the closest continuer theory, events in the future may change events in the past⁴⁵: if, at t_3 , *B* becomes identical with *A*, then, for instance, it becomes true of *B* that *B* (as *A*) came into existence at the time *A* came into existence, that *B* in its past underwent experiences that *A* underwent, and so on. And this may turn out to be true of *B* because of something that happens to *C*. So, for instance, suppose that at t_3 , *B'* (that is, Albert, Bernadette and George) realizes that all it has to do to be The Club is somehow to get rid of the Idaho contingent. It arranges for the group to set out in a leaky boat, which, as we learned in the last

⁴⁴ As before, Nozick is well aware of this implication. He writes: "It [does] seem[] strange that at a certain time, without any (physical) change taking place in it, the new [entity] could become the [original] when the [original] expires. However, once we have become used to the idea that whether *y* at t_2 is (identical with) *x* at t_1 does not depend only on the properties and relations of *x* and *y*, but depends also on whether there exists a *z* of a certain sort (which more closely continues *x*), then perhaps we can swallow this consequence as well" (Nozick 1981, 43-44). He adds: "This instantaneous movement of a person from one place to another does not violate special relativity's constraint[s]" (Nozick 1981, 660, note 13). Indeed, one might describe the situation less tendentiously by saying: "At t_1 , *A* was at location L, and at t_2 *A* was at location M, and *A* was nowhere in between." That is, one need not speak (except out of habit) of *A* having *moved* from L to M; one might take the consequence of the theory to be that things might appear at places without having *moved* to those places.

⁴⁵ Note that there is a trivial sense in which even ordinary theories of identity over time allow for the possibility that events in the future might change events in the past; anything that happens to *A*-later makes it true of *A*-earlier that *A*-earlier was an entity to which such events will happen. For instance, suppose that *A*-later undergoes *P* at t_2 . Then this event (which happens at t_2), makes it true of *A*-earlier that it will at some time be an entity which has undergone *P*. And this also makes it true of *A*-later that it is an entity with a predecessor (*A*-earlier) that has the property being-something-that-will-at-some-time-undergo-*P*. As will be clear from the example in the text, this is not the sort of case I am thinking of.

paragraph, sinks at t_4 . This allows the Indiana contingent to hoist the flag high and declare themselves The Club. But by doing so, the trio changes the past. They make it true of themselves that they are a club which has existed since t_1 (the time that the original six came together), and they do so with no substantial changes to themselves; everything that happens (in the ordinary sense of the term) happens to poor C⁴⁶.

Since one of the primary appeals of the closest continuer theory is its ability to account for our ordinary intuitions about identity over time, the fact that it produces such counterintuitive results is worrisome. If saying that the Istanbul group would be the Vienna Circle if the United States group had not existed means saying that identity over time is intransitive, that objects can move instantaneously from one place to another, and that events in the future can change events in the past, then perhaps it makes more sense to change some of our views about the Vienna Circle⁴⁷. For if that's what it means to say that to be something later is to be that thing's

⁴⁶ Mark Johnston describes another such case. Following Williams 1970b/1973, he supposes that there is a machine which has “read” A’s psychology into B’s body, and which is starting to “read” the psychology into C’s. The B-body person, who is walking around thinking: “I am A. I did not just come into existence,” sees the machine implanting the information in the C-body person, and realizes that if the process is completed, the C-body person will be a better continuer of A, and will, by the closest continuer theory, *be A**. So he shuts off the transmitting machine, and thereby *makes it true* that “he is A and has existed before the operation of the machine as A. But...surely our intuition is that the B-body person’s thought, ‘I am A. I did not just come into existence,’ is made true or false by what has happened up to and including the time at which that thought occurs. Surely no subsequent act by the B-body person can make this true or false” (Johnston 1991a, 68-69).

*See Nozick’s discussion of “Overlap” at Nozick 1981, 43-47 and 660 note 13.

⁴⁷ Note, however, that there is a complicated interplay among the counterintuitive implications that I have enumerated, such that one might be able to block the problem of changing the past by strengthening the force of non-transitivity. (That is, one might say that it simply isn’t true of B that things that are true of A are true of it, even if B is (the closest continuer of) A.)

closest continuer, then maybe we weren’t committed to such a view after all⁴⁸.

Still, the counterintuitive implications of the view might be tempered somewhat if we combine the closest continuer theory with something like van Inwagen’s or Parfit’s view, where we recognize that the claims we are making about identity are claims about common parlance, or about how we apply our concepts. So, for instance, if the question we are asking is what Parfit would call a conceptual question (see Section 3.5.2 above), then perhaps it is not so remarkable that our descriptions permit the sort of jumping about and intransitivity that we have been discussing in the last few pages. Again, if the question we are asking is not about what there is, but about how we speak about what there is, then perhaps it is not so remarkable that our descriptions permit us to attribute qualities to entities in the past on the basis of what happens to other entities in the future. But if there is some way of capturing what is true in the closest continuer theory without having to countenance such implications, then we will be better off.

3.7 THE PROPOSED DIAGNOSIS

Those who discuss the Ship of Theseus are concerned with describing a set of processes P1-Pn that come as close as possible to meeting the following three constraints:

- (a) *sufficiency*: each of the processes is sufficient for identity over time (if y is the result of subjecting x to some process Pi, then y is identical to x)
- (b) *necessity*: the processes are jointly necessary for identity over time (the question of whether y is identical to x can be settled by determining whether y is the result of subjecting x to some process Pi)

⁴⁸ Nozick’s solution seems to be to say that we could “get used to” such implications, once we recognize that they follow from the extrinsic determination of identity (see, for instance, Nozick 1981, 43-44). But since I think there is a better place to locate the extrinsic determination, I think we can avoid the need to get used to these consequences, while preserving the central insight.

- (c) *ascertainability*: the processes are relatively localized and surveyable by us (for instance, we can tell, without taking a full inventory of the universe, whether y is the result of subjecting x to some process P_i)

What the Ship of Theseus example brings out is that the three constraints come into tension with one another⁴⁹. Sets of processes that straightforwardly satisfy (a) and (c) run into trouble with (b) (for instance, a set of processes that concerns only cases where objects lose less than 1/16 of their matter), and sets that straightforwardly satisfy (a) and (b) run into trouble with (c) (for instance, a set of processes that requires full knowledge of the future location of all particles in order to determine whether a certain process obtains). (Those that satisfy (b) and (c) by giving up on (a) have changed the project, in that we are no longer given conditions for recognizing instances.) Faced with this dilemma, the solutions described in Sections 3.5 and 3.6 implicitly understand (c) in a certain way, and implicitly favor (b) over (c). In their efforts to state principles that cover all cases—including exceptional ones—they end up with principles that violate what I will argue are the correct sorts of epistemic requirements⁵⁰.

What I propose instead is that we restrict the scope of (b), and understand (c) in a way that allows us to build agent-intent and background patterns into our identity criteria. If we take as our primary task describing processes that provide identity conditions for normal cases (which themselves require attention to agent-

⁴⁹ One might think that this result requires an additional constraint, equivalent to:

- (d) *univocality*: no pair of processes gives divergent answers to identity-questions concerning a single object (if, if y is the result of subjecting x to some process P_i , then y is identical to x , then there is no other process P_j such that if z (not identical to y) is the result of subjecting x to P_j , then z is identical to x)

But (a) in conjunction with the logic of identity and an acceptance of common-sense ontology gives us (d), so it does not need to be stated separately.

⁵⁰ In the pages that precede (and in the pages that follow), no attempt has been made to give the full set of conditions that would be required to satisfy (b); all that has been attempted is a partial list of those directly related to Theseus-like cases.

intent and background patterns) then we can satisfy (a) and a restricted version of (b) without giving up on (c), properly understood. Resolving the trade-off between (b) and (c) in this way is preferable to the alternative. Restricting (b) and expanding (c) in the right ways leaves us with most of what we want in a theory of identity over time, whereas failing to do so leaves us with a theory which, though maximally wide-ranging, is minimally useful.

3.7.1 Some Very General Candidate Principles

The puzzle we have been discussing in this chapter is a consequence of subscribing to the following two very general principles, both of which are roughly true as candidate sufficient conditions for identity:

Disassembly Principle: objects can survive disassembly and subsequent reconstruction

Replacement Principle: objects can survive the gradual replacement of component parts over time

The question to be asked, then, is what makes the principles roughly *true* (that is: what do they get right about conditions of identity over time), and what makes them *roughly* true (that is: what do they get wrong about conditions of identity over time)?

Put as slogans, the disassembly principle says that same matter plus same form can allow gaps in form, whereas the replacement principle says that same form plus no gaps can allow different matter. This suggests the following analysis. In ascertaining whether an object x at t_1 is identical to an object y at t_3 we are concerned with relations among the following:

- (d) whether there is the same matter for x at t_1 and y at t_3 ;
- (e) whether there is the same form for x at t_1 and y at t_3 ;
- (f) whether there is the same form continuously for x/y at t_1 through t_3 ⁵¹.

⁵¹ Cases where we have the same matter at t_1 and t_3 but different matter at t_2 raise questions beyond the scope of discussion here.

So let us examine the relations among them. Cases where (d), (e) and (f) all receive an affirmative reply are governed by what we might call the:

Stability Principle: objects can survive being continuously composed of the same matter in the same arrangement over time

which is, for all intents and purposes, universally true. Cases where (e) is not satisfied are covered by what might be called the:

Distortion Principle: objects can survive (radical) changes in form over time

which is, generally speaking, not true (even where (d) is true). Cases where (d) and (e) but not (f) are answered affirmatively are governed by the Disassembly Principle; cases where (e) and (f) but not (d) are answered affirmatively are governed by the Disassembly/ Replacement Principle, both of which are, as mentioned above, approximately true. Cases with only (e) are governed by what we might call the Disassembly/ Replacement Principle, which is sufficient for identity under some but not all cases; and cases where none of the three is answered affirmatively are governed by what we might call the Distortion/Replacement Principle, which is sufficient for identity only rarely. In sum, the relation among the criteria is approximately captured by the following chart⁵²:

| (d) | (e) | (f) | Principle | Identity? |
|-----|-----|-----|-------------------------|------------------|
| Y | Y | Y | Stability | Yes |
| N | Y | Y | Replacement | Yes (generally) |
| Y | Y | N | Disassembly | Yes (generally) |
| N | Y | N | Disassembly/Replacement | Maybe |
| Y | N | N | Distortion | Rarely |
| N | N | N | Distortion/Replacement | Even more rarely |

⁵² If the answer to (e) is “N,” then the answer to (f) is “N,” hence the two missing rows.

3.7.2 Remarks on the Candidates

My goal in this sub-section is to look at the candidate conditions described above in light of normal cases, see what their application presupposes, and use this information to say a bit more about how the spirit of (c) should best be preserved.

Let's look first at Stability. Stability articulates an ideal sufficient condition for identity over time for objects that persist by spatio-temporal continuity⁵³. But as a condition for objects in this world, it is largely useless. Actual entities are responsive to their environments: they grow; they decay; they interact with other entities. Stability may describe what identity-conditions would look like in a static world, but criteria of identity over time are attempts to provide us with guidelines for navigating a world of flux. So while Stability may be a sufficient condition for identity in some abstract sense, even an ideal towards which other conditions aspire, it cannot serve as an intuitive basis for what we are after.

We turn next to the Disassembly Principle. As stated, the principle looks only at the beginning- and end-points of the process with which it is concerned: it says that objects produced by configuring the original component parts of some entity in a shape identical to the original entity can properly be said to be identical with the original entity, regardless of what happens in between. *Prima facie*, this seems plausible: what else, after all, could be relevant to the situation? Here we have the same stuff in the same shape; we must have the same entity.

⁵³ For other sorts of entities, this is not the case. Let us define the (grue-like) *Red House* as follows:

The *Red House* is that entity which is the weekday-portions of the White House, and the weekend-portions of the Kremlin*.

The Stability Principle is *not* a sufficient condition for identity for the Red House, but as I have remarked above, I am not attempting to offer a theory of identity which would cover such gerrymandered entities. (For further examples of entities of these sorts, see the opening chapters of Hirsch 1982.)

*In both cases, we'll use Greenwich Mean Time to determine the start of the day.

But even without the sorts of complications that a Theseus-like case introduces, the principle fails to articulate a sufficient condition for identity over time. Suppose when I was four years old I arranged three blocks into a pile and called that pile "Eiffel." The next morning, another child coincidentally arranges those three blocks in the same configuration, as a step in constructing a ten-block castle. It seems odd to say that Eiffel is back. Similarly, if a collection of matter once configured in some way comes, by pure chance, to be similarly configured at some point in the distant future, it seems odd to say that the entity once composed by that matter has come back into existence. So there are cases where the same matter in the same form might fail to constitute the same entity over time, either because it does not make sense to speak of that collection of matter in that form as constituting an entity, or because the causal and intentional relations that connect the two are insufficient. Disassembly, then, seems to require the following modification⁵⁴:

Disassembly Principle*: objects can survive disassembly and subsequent reconstruction given the right causal and intentional connections

Spelling out what those causal and intentional connections are is, of course, highly non-trivial. But what is important for our purposes is recognizing that this sort of modification must be introduced even to account for the most common of cases.

Unlike the Disassembly Principle, which looks only at the beginning- and end-states to ascertain whether the specified criteria have been met, the Replacement Principle is concerned with the sorts of *processes* to which an object can be subjected while maintaining identity. In particular, it is concerned with the process that involves the form-preserving gradual replacement of parts over time⁵⁵. What the principle says is that so long as the replacement is grad-

⁵⁴ A parallel modification will be required, for parallel reasons, for the other principles as well.

⁵⁵ I am concerned here, as before, with replacement of parts over and above that which occurs at the micro-structural level as a result of sub-molecular exchanges of particles.

ual enough, so long as the replacement respects the internal structures of the entity in question, gradual changes in matter might be made with no loss of identity over time⁵⁶.

This principle rests on the assumption that there are certain normal processes which entities of one or another kind can undergo, processes which respect certain patterns in the world which we pick out as significant, and for which we recognize certain sorts of transformations as normal and identity-preserving and others as abnormal and identity-destroying. The specific features of the pattern in question depend on what sort of entity we are considering⁵⁷; a lump of clay can survive changes that a statue cannot⁵⁸, a polygon can gain sides that would destroy a rectangle.

But what is crucial about this principle as far as our argument is concerned is that it brings out the ways in which the identity criteria for object A over time depends on all sorts of facts about objects other than A. Identity criteria over time depend on there being certain sorts of salient *patterns* by which we classify entities, and by which we make judgments about what sorts of changes entities of a certain sort can undergo⁵⁹. Identity in this sense is extrinsically determined; there would be no such thing as being the same ship unless *ship* were something an entity might be. So although it is true that identity for particular objects cannot be, as a matter of course, extrinsically determined (for the reasons discussed above in section 3.4), the *processes* by which we specify the conditions for identity may be, in the sense that they depend for their significance on the presence of certain background patterns.

⁵⁶ For a discussion of these matters as they relate to issues of personal identity, see Unger 1990, 123-125.

⁵⁷ Cf. Aristotle: "Some things are characterized by the mode of composition of their matter, e.g. the things formed by mixture, such as honey-water; and others by being bound together, e.g. a bundle; and others by being glued together, e.g. a book; and others by being nailed together, e.g. a casket; and others in more than one of these ways; and others by position, e.g. the threshold and the lintel (for these differ by being placed in a certain way); and others by time, e.g. dinner and breakfast; and others by place, e.g. the winds; and others by the affections proper to sensible things, e.g. hardness and softness, density and rarity, dryness and wetness; and some things by some of these qualities, others by them all, and in general some by excess and some by defect. Clearly the word 'is' has just as many meanings" (Aristotle, *Metaphysics* 1042^b 15-25 in Ackrill 1987).

⁵⁸ Assuming that the statue is essentially a statue.

⁵⁹ Cf. Hirsch's Sortal Rule, as presented in Section 3.6.1 above.

In this light, we might modify the Replacement principle as follows:

Replacement Principle*: objects can survive the gradual replacement of component parts over time given the right causal and intentional connections, and the right background patterns

Finally, a few words about the lessons to be learned from the general inapplicability of the Distortion Principle. The Distortion Principle concerns cases involving alteration of form but preservation of matter; the Replacement Principle, by contrast, concerns alteration of matter but preservation of form. That the Distortion Principle is widely restricted in its application whereas the Replacement Principle is not suggests that formal criteria generally take precedence over material ones in judgments of identity. This seems to give us reasons for thinking that the Replacement Principle (which involves preservation of form throughout) ought to trump the Disassembly Principle (which allows for distortion of form during the period between t_1 and t_3). Such a Trumping Principle would seem to solve the Theseus puzzle quite simply: the ship is to be identified with the continuously-repaired rather than the reconstructed ship.

3.7.3 A Messier Puzzle

That straight trumping can't serve the purpose for which it is intended is brought out by the following Theseus-like case. Suppose there is a pickle-barrel which I have inherited from my beloved grandfather. Each year, in order to insure that the barrel lasts a long time, I replace one of the wooden planks from which it is made with a new one. I hire the same company every year, and unbeknownst to me, they collect the planks I have given them and store them in their warehouse. One year, I replace the last of the original planks: as far as I'm concerned, I still have my grandfather's pickle barrel—the replacement of parts has been gradual, and I have done nothing to the barrel that he would not have done himself. Indeed, during the years that he owned the barrel, he himself replaced all of the planks several times over. He hired the same company every year, and unbeknownst to him, they collected all the planks he gave them over the years and stored them in their warehouse.

One day the son of the owner of the plank-collecting company comes up with the marvelous idea of making pickle barrels out of the planks languishing in the company's warehouses. Starting with the oldest planks, he assembles three pickle-barrels, each with a form exactly like the form of the pickle barrel that had belonged to my grandfather. But the similarity does not stop there. The first pickle barrel he assembles has exactly the same planks and exactly the same shape as the first pickle barrel my grandfather had ever owned, back in 1921 when Gendler's Groceries first opened its doors to the booming community of Chariton, Iowa. The second pickle barrel has exactly the same planks and exactly the same shape as my grandfather's pickle barrel in 1947, which he rolled out of the store and into his kitchen before the store was sold to some fancy investors from Des Moines. And the final pickle barrel he assembles has exactly the same planks and exactly the same shape as the pickle barrel that I inherited from my grandfather in 1977, when the preconditions required for the concept of inheritance to apply were, alas, made manifest.

Knowing that I am an aficionado of such things, the plank-collector's son comes to me with these three reconstructed vintage pickle barrels, and asks me whether I want to buy any of them. I do, and in the course of negotiations inquire as to the provenance of the materials used to construct them. As you might imagine, the consternation with which I greet his reply is profound.

The story I have told is just a double-barreled version of classic story of the Ship of Theseus, where instead of two, there are four plausible candidates for the title "my grandfather's pickle barrel":

- (i) the barrel I had in my house before the arrival of the plank-collector's son
- (ii) the barrel reconstructed from the planks that made up the barrel I originally inherited in 1977
- (iii) the barrel reconstructed from the planks that made up the barrel my grandfather rolled out from his store in 1947
- (iv) the barrel reconstructed from the planks that made up the pickle barrel my grandfather purchased in 1921.

What I mean by "candidate" is this: but for the existence of the other three, each of the pickle barrels I have described might plausibly have been called "my grandfather's pickle barrel."

The two principles of survival over time which together are sufficient to create this sort of puzzle case are the familiar principles we have been discussing all along: Reconstruction and Disassembly. Barrels (ii), (iii) and (iv) are candidates for identity with the original only if the Disassembly principle applies to them; barrels (i), (ii) and (iii)—the barrels that have components that were not part of the original barrel—are candidates for identity with it only if the Reconstruction principle applies to them. So barrel (i)'s candidacy makes appeal only to the latter principle, barrel (iv)'s candidacy only to the former, and barrels (ii)'s and (iii)'s candidacies to both. Suppose now that barrels (i) and (iv) are destroyed, and we are left only with barrels (ii) and (iii). It seems that we have no grounds for choosing between them. Straight trumping won't help us here: even if Replacement takes precedence over Disassembly where both apply, this will not allow us to adjudicate between (ii) and (iii). Moreover, it seems clear that there could be arbitrarily many such candidates of such a kind. So if the goal is to come up with principles that will allow us to determine identity in all possible cases, we will need to deal with cases far messier than the original Theseus case.

3.7.4 The Proposed Diagnosis

What I propose instead is that we take the lessons of 3.7.2 and extend them, modifying them in light of the following recognition. What causes problems in the Theseus and Theseus-like cases is a certain intertwining of the two principles. Just what sort of intertwining? In the case of Replacement, it involves doing something with the supposedly discarded *matter*; in the case of Disassembly, it involves doing something with the supposedly discarded *form*. So in both cases, if we attend to what happens with these “negative spaces,” we should be able to articulate principles that allow us to state sufficiency conditions without raising Theseus problems. Here are the modified principles:

Disassembly Principle:** objects can survive disassembly and subsequent reconstruction given the right causal/intentional connections given that we ask: what happens to the old places (form)? and the answer is: no new relevant matter goes into them (there is no substitution of parts)

That is to say: if x at t_1 and y at t_3 have the same matter M arranged in the same form F , and the right causal/intentional connections hold between them, then M need not have been arranged F -wise at times between t_1 and t_3 , so long as if this condition is taken to be sufficient, we do not also take it to be sufficient that some being composed of some other matter N arranged F -wise could also be sufficient for identity with x . More casually, if y results from disassembling and then reassembling x , then we want to say that y is identical with x only if, during the process of disassembling x , the parts removed from x were not replaced with other M -like parts⁶⁰.

Replacement Principle:** objects can survive the gradual replacement of component parts over time given the right causal/intentional connections and the right background patterns given that we ask: what happens to the old parts (matter)? And the answer is: no new relevant form is given to them (they have been discarded with regard to F -ness)

That is to say: if x at t_1 and y at t_3 have the same form F , and the right causal/intentional connections hold between them, and the right sorts of background conditions obtain, then y need not be composed of matter M , so long as this condition is taken to be sufficient, we do not also take it to be sufficient that being composed of M arranged F -wise could also be sufficient for identity with x . More casually, if y results from the gradual replacement of x 's parts over time, then we want to say that y is identical with x only if, during the process of disassembling x , the parts removed from x were discarded from F -use.

What these restrictions do is to give the right weight to (b) and (c). We are left with no definitive answer to the Theseus case, but that is as it should be. Let us return to the initial presentation of the puzzle in Sections 3.2 and 3.3. I suggested there that the story of the Ship of Theseus raises the puzzle of how a process that is ordinarily identity-preserving can in some cases be entity-creating. In particular, at least one of the two ordinarily identity-preserving processes we have been discussing in this chapter (Disassembly or Replacement) is, in the case of the Ship of Theseus, entity-creating.

⁶⁰ Note that if Replacement strongly trumps Disassembly, we will not need to include this restriction explicitly, as it will already be built into the relation between the principles.

But what should now be apparent is that to put the question this way is to presuppose that the process of part-replacement is the same process in the first version and the third version of the story, and that the process of disassembly-reassembly is the same process in the second and the third. And the obvious question to be asked is: why should we think that the process of part-replacement (or disassembly-reassembly) which is entity-preserving in the first (or second) version is the same process as that which takes place in the third version? That is, why should we think that processes are intrinsically determined in this sense? Why should what it is for a process to be a process of (say) disassembly and reassembly of the relevant sort be only a matter of what happens to the particular planks which are subject to being detached from one another, and then reattached to one another? Why might it not instead be that for such a process to be the relevant sort of process (that is, the same sort of process as the one described in the second version), certain things have to be true of other planks as well? In particular, perhaps to be the relevant sort of process, it must be true not only that the planks in question are detached from one another in a certain sort of gradual process, but also that no other planks are attached to these planks during the course of detachment? Or, to take the parallel case, perhaps to be the relevant sort of process, it must be true not only that the original planks are replaced sequentially by a set of planks of similar size and shape, arranged in the same manner, but also that the planks which are removed in the course of such a process be disposed of in some way which precludes reassembly⁶¹?

The Theseus case turns out not to be governed by principles to which we subscribe: what we are committed to are Disassembly** and Replacement**, and not their unmodified counterparts⁶². When a part is removed from an object and not replaced by a new

⁶¹ Cf. Smart (1972) and Smart (1973).

⁶² In content, this solution is quite similar to that proposed by Hirsch. He suggests that the Compositional Criterion be modified to include the clause that “there is no object z such that z comes into existence at a time t' between t_1 and t_2 ” which is itself a candidate for identity according to the Sortal Rule or the simple version of the Compositional Criterion (Hirsch 1982, 71). But the reasons he offers for subscribing to these conditions are different.

part, the part remains a candidate for later reconstitution of that same entity. This allows for identity-preserving disassembly and reassembly of watches and automobile engines of the sort we would like to countenance. But in other cases, when the process of disassembly takes place in a particular context, the subsequent reassembly of previous components will not produce a candidate for identity. That context is the context where a part that is removed is replaced by a new one.

In so doing, we help ourselves to the sort of constraint that is already in place when we think about identity over time. Just as we need to assume certain background patterns when we apply the Replacement principles to ordinary cases, so too do we assume certain background patterns when we apply the Disassembly principle. And these background patterns include certain assumptions about what sorts of processes the entity from which the parts are removed will undergo. When these processes are exceptional, as they are in the story of the Ship of Theseus, the principles cannot be applied straightforwardly.

CHAPTER 4

Personal Identity

4.1 INTRODUCTION: THE FACTS OF LIFE

Human beings come into existence through a well-known sequence of processes. Brought together by some sort of matchmaker, an egg fertilized by a sperm is implanted in the wall of a uterus, where it develops from a collection of a few cells into a progressively more complex entity. Roughly nine months later, the mother gives birth to a new human being, who then undergoes a long and intensive process of being fed and nurtured and interacted with by other human beings who have taken on the responsibility for the development of this one. After many years, the child reaches adulthood, becoming an autonomous agent who makes plans and commitments and decisions, and has experiences and relationships and ideas.

My goal in this chapter is to demonstrate that these well-known facts have received the wrong sort of attention in recent philosophical discussions concerning the nature and value of personal identity. It is not that they have been forgotten; no one, even in the throes of debate, really thinks that human beings sometimes come into existence through teletransportation or fission or brain-state-transfer. Rather, it is that such facts are deliberately treated as provincial truths: this may be the way we get persons off in our little corner of possibility space, but simply considering specimens from our village will not tell us everything we would want to know about what sort of things persons are. Many of the features that all of us share are just local color. And we wouldn't want to be

such yokels as to mistake these coincidental features of persons-as-they-happen-to-be for the constitutive features of persons-as-they-truly-are¹.

My aim in the succeeding pages is to highlight a danger that plagues such attempts to overcome provincialism². Merely contingent constraints on the way things happen to be may play central roles in our categories of evaluation; to the extent that this is so, cases where we “imagine away” such constraints will be systematically uninformative as guides to the concepts we set out to explore. The fact that two features can be conceptually separated in the sense that we can imagine the one obtaining without the other in some particular case is not sufficient to establish that those two features are conceptually distinct in the sense that we can know how we would evaluate the one if it obtained without the other systematically. Even if we are aware that the two features need not coincide in all possible cases, the fact that they coincide in all actual cases may mean that there is no ascertainable fact of the matter about how we would or should respond to either in isolation.

¹ Cf. Shoemaker: “What Mackie and Perry have done is to indicate how personal identity (or copersonality, or psychological unity) are realized in us, i.e. in members of our own species. And this does not answer the question ‘What does personal identity consist in?’ at the level of abstractness at which we want it answered” (Shoemaker and Swinburne 1984, 127). Or again Unger: “[W]hy not stick only to actual cases?...The reason is that this extremely conservative methodology is apt to incur great costs...In attempting to ascribe beliefs to ourselves on the basis of quite limited data, we might wrongly describe our own attitudes” (Unger 1990, 11). Or Nozick: “We...are not so tied to our bodies that we find it impossible to imagine coming to inhabit another. We do not conceive of ourselves as (merely) our particular bodies, as inextricably tied to them” (Nozick 1981, 30).

² These attempts involve appeal to imaginary cases which, like experiments, are supposed to help us compensate for the often arbitrary way in which we tend to come upon information in the world. Experiments isolate sets of phenomena so that the relations among them are made manifest, and thereby reveal in an epistemically accessible way patterns that, in some sense, were already out there to be seen. In a parallel fashion, thought experiments are supposed to help us distinguish relevant from irrelevant features in actual cases by making manifest their relations in cleaner, non-actual, cases.

My goal in this chapter is to demonstrate the truth of a particular instance of this general claim. What I will argue is that a certain widely-discussed argument of Derek Parfit’s concerning the importance of personal identity is unsuccessful because it ignores the sort of facts I described in the first paragraph—facts that describe the contingent concomitance of a number of features which can, in imaginary cases, be conceptually separated. Parfit tries to show that what ought rationally to matter to us when we care about survival and future well-being is not that we ourselves survive, but only that someone exist who is psychologically continuous with us in the right sort of ways. He describes an imaginary case where it seems clear that identity would not be what matters in this way, and argues that if we are consistent in our commitments, then we ought to conclude that even in actual cases, identity is not what matters³.

What I try to show is that a strategy that on its surface seems totally untenable can in fact be used to block Parfit’s conclusion. I argue that what explains the rationality of prudential concern is always identity, while accepting that in some cases rational prudential concern might hold in the absence of identity. I do so by showing how the feature that explains or justifies or makes rational a relation can be a different feature from the one that underpins it as a necessary condition. This allows us to accept that there

³ Parfit writes: “By considering these cases, we discover what we believe to be involved in our continued existence...Though our beliefs are revealed most clearly when we consider imaginary cases, these beliefs also cover actual cases, and our own lives” (Parfit 1984/1987, 200). For discussions of related methodological questions, see, among others: Aune 1994; Baillie 1993a, esp. 200-205; Baillie 1993b; Baillie 1996; Beck 1992; Cockburn 1991, *passim*; *Ethics* 96 (July 1986), *passim*; Donagan 1990; Gale 1991, 299-303; Gendler 1998a; Haksar 1991, 149-155; Hertzberg 1991, 153-155; Johnston 1987a, esp. 60-69, 80-83, 1989a, 1989b, 1992, 1997; Kolak 1993; Martin 1993, 1995; Oderberg 1993, 32-36; Quine 1972, 490; Rovane 1990 and 1998, *passim*; Snowdon 1991; Sosa 1990; Unger 1990, Chapter 1, esp. 7-15, 27-35; White 1989; Wiggins 1980, Chapter 6, esp. 169-175, 221; Wilkes 1988, Chapter 1. My own ideas on personal identity have been strongly influenced by the writings of Mark Johnston, especially his discussion of Parfit in “Reasons and Reductionism” (Johnston 1992) and “Human Concerns without Superlative Selves” (Johnston 1997). For those who know these works, my debts to Johnston should be obvious.

might be cases of the sort Parfit describes, without taking them to have the implications that he takes them to have.

After offering a short discussion of the context into which Parfit's case fits (Section 4.2.1) and of the ways in which fission cases are and are not relevant to us (4.2.2), I move to the central argument of the text. I begin by presenting Parfit's argument (4.3.1) and identifying three crucial assumptions that underlie his reasoning:

The Intrinsicness Principle (for M): The relation that matters for rational prudential concern—M—is an intrinsic relation.

The Sufficiency Principle: A's prudential concern for B is rational (RPC) if the relation which matters for rational prudential concern (M) holds between A and B.

The Necessity Principle: A's prudential concern for B is rational (RPC) only if the relation which matters for rational prudential concern (M) holds between A and B.

Together, these principles generate a particular internalist commitment: they imply that there is a certain relation (M) which is both necessary and sufficient for rational prudential concern (of A for B), a relation whose obtaining depends only on facts about A and B. My goal in the remainder of the chapter is to show that this commitment is reasonable on one understanding of what M involves, but not on another, and to show that Parfit's argument rests on an equivocation between the two senses.

The first sense in which we might be interested in M is this: we might be interested in finding the relation which represents the Greatest Common Factor among cases where rational prudential concern obtains. But there is a second sense in which we might be interested in M: we might be interested in finding the relation which explains why rational prudential concern obtains. Of the three principles just enumerated (the Intrinsicness, Sufficiency and Necessity Principles), all three are true when we are interested in M in the first—Greatest Common Factor—sense⁴. But if we are

⁴ In the interests of simplicity, I am assuming that M is believed to obtain iff M obtains (this bears on the truth of Sufficiency). I am also assuming that disjunctions of intrinsic relations are themselves intrinsic (this bears on the truth of the Intrinsicness Principle if we allow that M may be disjunctive.) (Thanks to John Hawthorne for pointing out the need for these caveats.)

interested in M in the second—explanatory—sense, neither the Intrinsicness Principle nor the Necessity Principle is true. Part of the apparent success of Parfit's argument can be attributed to a failure to make this distinction. I address this in Sections 4.3.2 through 4.3.5.

But while making this distinction allows us to see a certain conceptual possibility, it doesn't seem to capture much about why the argument feels so convincing. My aim in the remainder of the chapter is to diagnose the source of its apparent persuasiveness. I try to show that the reasoning on which Parfit depends is a special case of a seemingly undeniable principle of rationality, what Mill called the *Method of Agreement*. The Method of Agreement says roughly that if there is a single feature present in all cases where a given phenomenon obtains, then it is that feature which explains the obtaining of the phenomenon. But while such process-of-elimination reasoning is legitimate in cases of scientific analysis, it cannot be used where the issue at hand concerns explanation of value. My goal in Sections 4.4 and 4.5 is to motivate and defend this view.

4.2 SETTING THE STAGE

4.2.1 A Context for Parfit's Argument

Over the last forty years, a sizable literature has developed describing imaginary cases meant to illuminate questions concerning the metaphysics of personal identity and the rational grounding of our relations between our present and our future selves^{5/6}. Such cases are of two kinds. The first kind exploit the assumption that two

⁵ Apparently, such cases were also widely discussed in the eighteenth and nineteenth centuries; see “Fission Examples in the Eighteenth and Early Nineteenth Century Personal Identity Debate” (Martin *et al* 1998).

⁶ To the extent that they involve the contemplation of scenarios in which a human being is brought into existence who may be identical to no previously existing human being, these cases violate the predictable natural sequence of events that I reminded us of in the first paragraph. But unlike Greek myths and other people's religions, these cases are not invoked as poetic expressions of our deepest desires: that we should overcome our bodily engagement, that we should be able to create objects at will, that simply having enough knowledge about a person—knowing what she is like down to the last micromolecule—should be sufficient to bring her back. These are cases we consider, not as the production of some odd tribe of...

Continued on next page

structurally identical individuals composed of qualitatively identical matter will manifest the same qualitative characteristics, both physically and in terms of non-physical states. Such cases describe situations where some individual (who may share none of the same matter as the ordinarily-produced person she comes to resemble) comes into existence by some reliable or unreliable process that produces an individual in all ways indistinguishable from an ordinary human being. Often these cases involve descriptions of procedures, such as teletransportation or brain zapping, which involve the microstructural reconfiguration of found matter in order to produce macrostructural results. Since in all actual instances, substantive and qualitative criteria of individuation and valuation coincide, these cases are meant to tease apart which of the features in question is actually doing the grounding work.

Other cases, following the assumption that whatever produces consciousness can be localized to some fairly small part of the human organism, in particular the brain, involve thinking about what would happen if the brain associated with some human being were transplanted into some other body, either intact, or, with the additional assumption that only part of the brain is necessary to support the array of psychological characteristics in question, in some attenuated form. The first of these is generally referred to as brain-transplant, the second as fission. These cases are meant to tease apart the extent to which the commitment to identity and value is bodily based, and the extent to which we require only psychological and qualitative similarity, along with (perhaps) some of the same matter and (perhaps) the right kind of cause.

My focus in this chapter will be on cases of fission. Before moving on to Parfit's particular example, let me make some brief remarks about what we might expect cases of this sort to show.

4.2.2 What Fission Might Show

Assuming, as most moderns do, that psychological characteristics supervene (at least weakly) on physical microstructure, it follows that persons are, at base, complex physical objects. It is a general fact about complex physical objects that their identity conditions seem to allow for multiple-candidacy; this is a direct consequence

philosophers, but as examples about which the convergence of opinions is going to reveal our deepest commitments and beliefs about our nature as human beings. (See note 3.)

of the relation that underpins their identity allowing for preservation of identity under conditions of growth, change, or mutilation. As a result of this unavoidable leniency, there are at least two sorts of cases in which a process that is ordinarily identity-preserving may instead be entity-creating⁷. The first sort—Ship of Theseus cases—play off the fact that two distinct sorts of procedures are, in ordinary cases, sufficient for the continued existence of an entity; the second sort—fission cases—play off the fact that an entity might continue to exist despite the sudden loss of half or more of its matter. Having discussed cases of the former type at some length in the previous chapter, I will restrict my remarks in this chapter to cases second kind.

The series of assumptions that together seem to render consideration of fission cases mandatory can be put in terms of these statements, which might be characterized as “we would consider the following sort of being to be relevantly like us” statements.

We would consider the following sort of being to be relevantly like us:

⁷ There are two other sorts of cases which might be candidates for this description, or at least for the description: “a process that is often identity-preserving which may instead be entity-creating.” One would be cases involving creation of an exact qualitative duplicate; the other would be cases involving instantaneous loss of a bulk of matter. Both sorts of processes are sometimes considered person-preserving in circumstances where there is no equally good or better competitor: the first in teletransportation, the second in survival-as-brain.

I think cases of this second sort can be assimilated either to cases like Theseus cases, which I have discussed in Chapter 3, or to cases like fission cases, where the same criterion is equally well-met twice over by some bit of organized substance. So I don't think these cases raise genuinely new issues.

Cases of the first sort (teletransportation cases) do present a genuinely new kind of case, but here, I think, we have overstepped the bounds of identity. That qualitative features alone (or qualitative features constrained by some suitably replicable causal process) could serve as the criterion for numerical identity of entities of a certain kind would make sense only in a context where there is some sort of nomological near-guarantee of uniqueness (which is exactly what the supposition of teletransportation denies), and even here, it seems to me highly suspect (for reasons discussed by, among others, Wiggins 1980, 208; Williams 1956/1973, 1960/1973 and 1970a/1973.)

- (i) an organism whose psychological activity—knowledge, belief, memories, intentions, tacit motivations, etc.—supervenes on the physical structure of its brain. [Note: this need not be its brain; it could be anything (compact and physical), in which case replace “brain” with that throughout.]
- (ii) an organism meeting condition (i) from which it is technically possible to remove the brain surgically in such a way that, suitably resituated, the brain would support the same psychological activity before and after the surgery.
- (iii) an organism meeting condition (ii) such that when the brain is surgically implanted in a body other than the original, its relation to the new body with regard to sensori-motor control is qualitatively like its relation to its original body.
- (iv) an organism meeting condition (iii) whose brain is such that all features structurally relevant to its psychological functioning are realized in duplicate.
- (v) an organism meeting condition (iv) where the duplication would be such as to allow the brain to be surgically divided into two distinct sub-brains, each meeting the removal and transplant conditions as described in (ii) and (iii).

Defenders of fission contend that the depictions in (i)-(iii) describe ways that persons might actually (turn out to) be, so they pose no conceptual problem for the example. We would (and should), they think, consider those sorts of beings to be relevantly like us. As far as defenders of fission are concerned, the only potential problems lie with (v), where the depiction might be problematic for two distinct reasons⁸. The first, “deep,” reason would be that a single consciousness could never be divided in such a way

⁸ Cf. Parfit 1984/1987, 255.

as to produce two distinct streams; the second, “shallow,” reason would be that the lower brain of human beings, as it is actually constituted, could not be divided without rendering it non-functional. Both of these problems are treated as undamaging by defenders of fission. Because of empirical data suggesting actual instances exist of divided upper hemispheres resulting in divided consciousness⁹, the first reason has been widely rejected¹⁰. And though it has been a subject of great controversy in the literature, the second reason is standardly dismissed on the grounds that this fact about human beings plays no role in our central beliefs about the nature of persons¹¹.

Contra the defenders of fission, I think the problems begin before we reach (v). It is important to notice that there is a great deal of arbitrariness in treating certain presumably contingent features of human beings as fixed, and others as malleable. Certainly it seems to be the same sort of biological fact that the basis for our psychological activity is localized in the brain (i), and that the basis for our linguistic ability is localized in one of the hemispheres (the denial of (iv)). It is far from trivial to suppose that all psychological characteristics might be realized in the brain twice over (iv), and far from trivial to suppose that the brain stem and spinal cord would enjoy the same sort of duplicatory structure (v). It would be at least as plausible to suppose that the source of psychological activity might be totally non-localized (the denial of (ii)), so that the brain transplant with which the fission case begins would produce nothing but confusion (the denial of (iii)). If one were to argue that a non-localized center of psychological activity would run counter to evolutionary laws, which suggest that the center of

⁹ Cf. Nagel 1971, which introduced the research in question to the community of analytic philosophers. For more recent discussions, see works listed in the bibliography “Personal Identity and Identity”. See Bibliography 6.3 page 194.

¹⁰ I remain unconvinced, both by the evidence Parfit himself adduces, and by the other articles I have read on the subject, that these cases describe a genuine division of consciousness of the sort necessary to support his conclusion. Cf. Robinson 1988, esp. 325; Putnam 1981, 89-92. But since I think Parfit’s argument is problematic for other reasons, I will concede this point for the sake of argument.

¹¹ Both Robinson 1988 and Wilkes 1988 disagree, as does Wiggins 1980; but see Johnston 1989a, 376-377; Garrett 1990, 178-180 for replies.

intelligence should be located in a reasonably compact, well-protected area of the body, one would face equal difficulties explaining how duplication of all psychological features (or even of a significant proportion of high-level psychological features) could be countenanced on evolutionary grounds¹². Such mechanisms are extraordinarily expensive, far beyond the purchasing power of a satisficing spender like natural selection. It seems that (iv) is already counter-nomological, and that (v) is even more so.

But for the sake of argument, let us grant to the defender of fission that we might be creatures of the kind described in (v). What I will argue below is that even if it should turn out that we are such creatures, we would not consider the creature described in (v) to be relevantly like us in the sense that Parfit requires. That is, even if it turns out that we are creatures for whom it is not just metaphysically possible—but also medically possible—that we should divide, we should not conclude that identity is not what matters for prudential concern. What matters for prudential concern might be connected to certain contingent facts about the way things (merely) happen to be.

4.3 THE ARGUMENT AND ITS CRUCIAL ASSUMPTIONS

4.3.1 Parfit's Fission Argument

Parfit's familiar fission argument can be reconstructed as follows¹³. We are to imagine three triplets who are involved in an accident in which the body of one—call him Brainy—is fatally injured, while the brains of his two brothers are totally destroyed. Brainy is such

¹² This is not the same thing as suggesting (as has empirically been demonstrated) that certain memory traces are stored at multiple locations in the brain.

¹³ The version of the argument that I present below is a modification of that found in Parfit 1994a. In *Reasons and Persons*, Parfit presents the case as follows:

My Division. My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each 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. And he has a body that is very like mine (Parfit 1984/1987, 254-5)

Continued on next page

that the physical bases for his psychological characteristics happen to be realized in duplicate, one complete set in each lobe. Following the accident, doctors divide his brain in half, and transplant the two hemispheres into the bodies of the two brothers.

In the first scenario, which we might call the “single-transfer case,” only the left transplant takes, and the right transplant is destroyed. The resulting individual, whom we will call Lefty¹⁴, has all of Brainy's memories and psychological states and a body almost indistinguishable from the one that Brainy had before the accident. Parfit holds, and for the sake of argument we will grant him that:

(1) In the single-transfer case, Lefty is Brainy¹⁵.

Parfit presents the vast majority of his examples in a first-person form (see Parfit 1984/1987, 199ff). Although this convention certainly provides convenience of locution, I think it also reflects an attempt to bring the reader to imagine the cases “from the inside.” My decision to use the third-person in discussing this case is on stylistic grounds alone. For a discussion of the issues involved in first-person as opposed to third-person thought experiments, see Shoemaker 1994. See also Madell 1991, esp. pp. 128-129.

¹⁴ Note that “Lefty” is not a name; it is an abbreviation for the description “the individual who has Brainy's original left hemisphere.” For a discussion of the sorts of confusion that arise from failure to recognize this, see Shoemaker 1984, 116-118.

¹⁵ Parfit takes it as common ground both among supporters of the physical criterion of personal identity and among supporters of the psychological criterion of personal identity that person P₁ at t₁ is identical to person P₂ at t₂ if:

- (i) P₂ has enough of P₁'s brain to preserve psychological continuity*,
- (ii) P₂ has all of P₁'s psychological characteristics, and
- (iii) the brain in question has been continuously and properly functional from t₁ to t₂
- (iv) there are no other (equally good) candidates for identity with P₁

This position is rejected by Williams (see Williams 1973, *passim*), and more recently, by Thomson 1997 and Olson 1997a, who endorse what might be called a “bodily criterion.” But as Parfit has pointed out (personal communication), his fission argument might be made without appeal to a brain transplant. In this version, the one-sided case would involve Brainy's survival under conditions in which half his brain and body are destroyed, but the remainder (sufficient to support full psycho-

Continued on next page

Given this, along with the principle that:

- (2) If A is identical to B, then the relation which matters for rational prudential concern (M) holds between A and B.

it follows trivially that:

- (3) In the single-transfer case, M holds between Brainy and Lefty.

This process of reasoning is intended to establish a base case for an argument from parity. The parity argument concerns a second scenario, which we might call the “double-transfer case.” In the double-transfer case, both transplants are successful. Each of the two resulting individuals—we’ll call them Lefty and Righty—has all of Brainy’s memories and psychological states and a body almost indistinguishable from the one that Brainy had before the accident.

logical continuity) is kept functional by means of appropriate technology. The two-sided case would involve a corresponding division of Brainy’s brain and body into two such medically viable psychological continuers.

I am not convinced that this reformulation would be sufficient to satisfy his opponents. The reason half-brain survival is sufficient to guarantee identity for holders of a psychological or brain criterion of identity is because (*ex hypothesis*) it is sufficient to support full psychological continuity. But it is not clear what the analogous critical mass would be in the case of bodily survival; that is, as sufficient brain-survival is to psychological continuity, so sufficient body-survival is to what? (Medical viability does not seem to be a clear enough criterion, since there is no “whole” which the partial-body can nonetheless sustain.) In any case, this reformulation does not help Parfit in countering closest continuers or multiple-occupancy type objections. (Indeed, it may even weaken his case against the latter, in that it seems to lend support to the view that in any one person, there are two subpersons waiting to get out.)

One further remark: note that if one were to “ask” the “body” (rather than the “mind”) about the brain-transplant case, it would surely prefer bodily (i.e. genetic) to mental survival, and that even if the person who results from the brain transplant is identical with the person whose brain was transplanted, the children to whom he or she will give birth will inherit genetic information from the recipient body. Darwin wins, not Lamarck: if you put A’s brain in B’s body, the resulting person will give birth to B’s children.

*As evidence for the widespread acceptance of this view, consider the following “Incredible Brain Fact” from the back cover of the box enclosing “The Incredible Growing Brain” (a children’s toy made in Taiwan): “We would not be ourselves if our brain was transplanted [sic].” (I thank Teresa Robertson for the citation.)

Parfit points out that:

- (4) Brainy’s relation to Lefty is intrinsically the same in the single- and double-transfer cases.

From (3) and (4) plus:

- (5) The Intrinsicness Principle (for M): The relation which matters for rational prudential concern—M—is an intrinsic relation¹⁶.

It follows that:

- (6) The Parity Result (for M): If M holds between Brainy and Lefty in the single-transfer case, then M holds between Brainy and Lefty in the double-transfer case.

¹⁶ What makes a relation intrinsic as opposed to extrinsic is a vexed question, and I have little new to add in the way of attempted answers. [For a recent discussion see Langton and Lewis 1998.] For the purposes of this discussion, however, the following intuitive characterization will be sufficient: a relation is intrinsic if it supervenes on facts about the relata, and extrinsic if it supervenes on facts about objects distinct from the relata. Cf. the following formulations of the principle for cases of identity, which Wiggins and Noonan endorse, and Nozick and Parfit deny. Wiggins’s Only-a-and-b Principle: “For a relation R to be constitutive of the identity of a and b, a’s having R to b must be such that objects distinct from a and b are irrelevant to whether a has R to b” (Wiggins 1980, 96); Noonan’s Only-x-and-y Principle, defined (for personal identity) as: “Whether a certain later person P₂ is identical with a certain earlier person P₁ can depend only on facts about P₁ and P₂ and the intrinsic relationships between them; no facts about individuals other than P₁ and P₂ can be relevant to whether P₁ is the same person as P₂” (Noonan 1989, 16); Nozick’s Relevance Principle: “If x at time t₁ is the same individual as y at a later time t₂, this can depend only upon facts about x, y, and the relationships between them. no fact about any other existing thing is relevant to (deciding) whether x at t₁ is (part of the same continuing individual as) y at t₂” (Nozick 1981, 31); and Parfit’s Intrinsicness of Personal Identity Principle: If some future person will be me, that fact must depend only on the intrinsic features of the relation between me now and that future person. It cannot depend on whether the same relation holds between me now and some other future person” (Parfit 1984/1987, 267).

And from (3) and (6) follow:

- (7) M holds between Brainy and Lefty in the double-transfer case.

Now let us grant Parfit that Lefty and Righty are not the same person. (After all, the two occupy distinct spatial locations, undergo different experiences, and have no particular causal effect on one another.) If Lefty and Righty are different people, and Brainy is a single person¹⁷, then Lefty and Righty cannot both be identical to Brainy. So:

- (8) In the double-transfer case, Lefty is not Brainy¹⁸.

But since (7) tells us that the relation which matters for rational prudential concern holds between them, it follows that:

- (9) In the double-transfer case, the relation which matters for rational prudential concern is not identity.

And if whatever it is that matters for rational prudential concern holds between Brainy and Lefty in the double-transfer case, and whatever that is is not identity, then how can it be identity that matters for rational prudential concern in the single-transfer case? Consistency dictates that:

- (10) In the single-transfer case, the relation that matters for rational prudential concern is not identity.

¹⁷ For challenges, cf. Lewis 1976; Lewis 1983b; Noonan 1989, 164-168, 197-198; Mills 1993; Robinson 1985.

¹⁸ Strictly speaking, of course, additional assumptions are required for this step. The simplest argument would make a straightforward appeal to some principle of sufficient reason as follows: Since by stipulation there is no relevant difference between Brainy's left and right lobes, there is no reason that Brainy would be identical with Lefty as opposed to Righty, or with Righty as opposed to Lefty; and since he cannot be identical with both, then he must be identical with neither. More elaborately, one might proceed by describing two versions of the single-transfer case, the first with the right lobe, the second with the left; from this and the fact that Brainy is not identical to both Lefty and Righty, it follows that Brainy in the double-transfer case bears an intrinsic relation like that borne in cases of identity towards at least one person with whom he is not identical.

and that:

- (11) The Unimportance of Identity Conclusion: Even in ordinary cases, the relation that matters for rational prudential concern is not identity.

4.3.2 Four Crucial Distinctions

In evaluating Parfit's argument, four crucial distinctions need to be made. While this machinery may seem excessive when presented all at once, its importance will become apparent in the remainder of Section 4.3. (A visual representation of the relations among these distinctions can be found at the end of Section 4.3.3.) It is by appealing to these distinctions that I will try to articulate the ways in which I think that Parfit's argument goes subtly awry, and to identify the ways in which its going awry can be traced to its failure to respect the dependence of our concepts on the way things happen to be.

I thus begin by distinguishing between two different states of affairs:

- (a) M(A,B): the relation which matters for rational prudential concern holding between A and B

- (b) RPC (A,B): A's prudential concern for B being rational

and between two ways in which these states of affairs might be associated:

- (c) The Sufficiency Principle: A's prudential concern for B is rational if the relation which matters for rational prudential concern holds between A and B; in shorthand: $M(A,B) \rightarrow RPC(A,B)$.

- (d) The Necessity Principle: A's prudential concern for B is rational (RPC) only if the relation which matters for rational prudential concern (M) holds between A and B; in shorthand: $RPC(A,B) \rightarrow M(A,B)$.

The first distinction—between M and RPC—reminds us that there are two sorts of things in which we might be interested. We might be interested in whether A's prudential concern for B is rational—that is, whether $\text{RPC}(A,B)$. Or we might be interested in whether a certain relation which I am calling M holds between A and B—that is, whether $M(A,B)$. Whether these two interests amount to the same thing will depend on how M and RPC are related. M might be necessary for RPC, or sufficient for RPC, or neither, or both. If it is necessary for RPC, then the Necessity Principle is true; and if it is sufficient for RPC, then the Sufficiency Principle is true.

But in order to say whether I think the Necessity or Sufficiency Principles are true, I need to give labels to two different ways of understanding M, and to two different ways of understanding RPC. In the case of M, “the relation which matters for rational prudential concern” might be understood in at least two ways:

- (e) The Greatest Common Factor Reading (for M): the relation which matters for rational prudential concern means: whatever relation it is that has to be present in any case where rational prudential concern obtains.
- (f) The Explanatory Reading (for M): the relation which matters for rational prudential concern means: whatever relation it is that explains the rationality of prudential concern obtaining in cases where rational prudential concern obtains.

And in the case of RPC, “A's prudential concern for B being rational” might again be understood in two ways¹⁹:

- (g) The Permission Reading (for RPC): A's prudential concern for B is rational means: the norms of rationality permit A to be prudentially concerned for B.

¹⁹ I thank Carol Rovane for pressing me on the need to make this distinction.

- (h) The Requirement Reading (for RPC): A's prudential concern for B is rational means: the norms of rationality require A to be prudentially concerned for B.

So in the case of M, we might be interested in the relation that has to be present in all cases where RPC obtains. Or we might be interested in the relation that explains the obtaining of RPC. And in the case of RPC, we might be interested in the question of whether prudential concern is rationally required, or whether it is rationally permitted.

4.3.3 Comments on these Distinctions

In the interests of clarity, I now need to make a few brief comments on (c), (g) and (h), and longer comments on (d), (e) and (f):

Comments on (c): Throughout the chapter, the Sufficiency Principle (c) should be understood as holding by definition: on whichever reading we take, wherever the relation which matters for rational prudential concern holds between A and B, A's prudential concern for B is rational.

Comments on (g) and (h): I will be concerned only with the Permission Reading (g) and not the Requirement Reading (h) of RPC; A's prudential concern for B is rational should be understood as meaning: the norms of rationality permit A to be prudentially concerned for B²⁰.

Comments on (d), (e), and (f): I ask the reader to note the following. On the Greatest Common Factor Reading (e), the Necessity Principle (e) is trivially true. If the relation that matters for rational prudential concern means: whatever relation it is that has to be present in any case where rational prudential concern obtains, then clearly A's prudential concern for B is rational (RPC) only if the relation which matters for rational prudential concern (M) holds between A and B²¹. But on the Explanatory Reading (f), the connection is not so clear. If the relation that matters for rational prudential concern means: whatever relation it is that explains the rationality of prudential concern obtaining in cases

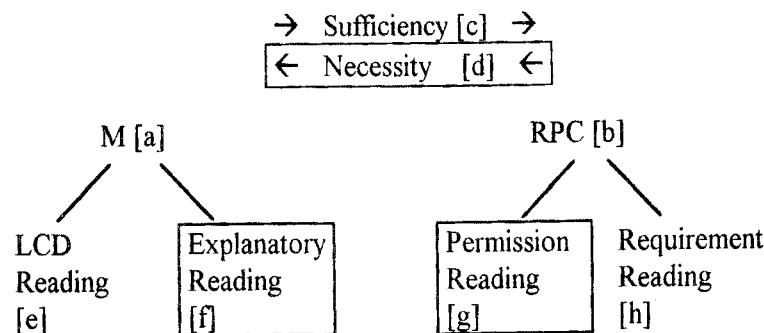
²⁰ It may be that the relation between prudential concern and rationality is such that prudential concern's being rationally permitted is tantamount to its being rationally required. My argument below is agnostic on this question.

²¹ See footnote 4.

where rational prudential concern obtains, it does not follow—or so I will argue in section 4.5—that A's prudential concern for B is rational only if M holds between them.

Moreover, it turns out that for Parfit's purposes, the Greatest Common Factor reading (e) is not adequate. For on this reading, the Unimportance of Identity Conclusion (11) says merely that even when we are thinking about ordinary cases, there may be other cases where there is rational prudential concern without identity. But Parfit needs much more than this; if we respond to the case in the way he expects, he thinks we should change our views about what underpins our prudential concern for our future selves. All the Greatest Common Factor reading (e) shows is that identity need not be present in every case where prudential concern is rational. This is something far weaker than what Parfit means when he says "identity is not what matters;" he needs to show that identity is not what matters in the explanatory sense (f), that identity is not what justifies prudential concern.

In sum: In considering the connection between RPC (b) and M (a) we are concerned with the Explanatory Reading (f) of the latter, and the Permission Reading (g) of the former. And we are interested in the question of whether on these readings, the Necessity Principle (d) is true.



In jargonless prose, the issue can be put as follows: must the relation which explains rational prudential concern hold between A and B in all cases where A's prudential concern for B is rationally permitted? What I will argue below is that the answer to this question is "no."

4.3.4 The Intrinsicness Premise

With these distinctions in place, let us return to Parfit's argument. The crucial premise in Parfit's argument is, of course, (5):

The Intrinsicness Principle (for M): The relation which matters for rational prudential concern—M—is an intrinsic relation.

It is the Intrinsicness Principle (for M) which allows Parfit to derive the Parity Result (for M), and it is the Parity Result (for M) which allows him to derive the Unimportance of Identity Conclusion.

I have no grounds for rejecting the Intrinsicness Principle (for M) where M is understood according to the Greatest Common Factor Reading (e). But I deny the truth of the Intrinsicness Principle (for M) where M is understood according to the Explanatory Reading (f). Whether the relation which explains the rationality of prudential concern holds between A and B does not depend only on facts about A and B; it also depends on facts about entities other than A and B. (I will argue for this in the sections that follow.) If this is so, then the relation which explains the rationality of prudential concern could be—as I claim it is—identity.

4.3.5 Summary

Parfit's goal is to show that "personal identity is not what matters"²² or, in our terms, that the relation which matters for rational prudential concern is not identity. He seeks to do this by presenting a case—the double-transfer case—where what matters for rational prudential concern is present, but identity is absent. And if the former can obtain without the latter, then identity cannot be what matters for rational prudential concern.

I suggested a number of distinctions which need to be drawn in evaluating the force of this argument. I pointed out that we need to separate questions about the relation I have been calling RPC (the relation that holds when A's prudential concern for B is ration-

²² Cf. Parfit 1984/1987, 255.

al) from questions about the relation which I have been calling M (the relation that matters for rational prudential concern). And I pointed out that we need to distinguish two ways of understanding each of these relations: RPC can be understood on either a Permission or a Requirement Reading; and M can be understood on either a Greatest Common Factor or an Explanatory Reading. I suggested that if Parfit wishes to establish the conclusion he does, then we need to understand RPC according to the Permission Reading, and M according to the Explanatory Reading.

With these distinctions in place, I suggested, we can diagnose part of the force of Parfit's argument. The principle which I have been calling the Intrinsicness Principle is arguably true for RPC. And I am prepared to accept that there is a case—the double-transfer case—where RPC obtains without identity. But this is not sufficient to establish that M obtains without identity.

4.4 TWO UNSUCCESSFUL STRATEGIES

4.4.1 An Unsuccessful Attack on the Intrinsicness Premise

One strategy for defending the position I have been advocating would be to challenge Parfit's argument by denying the Intrinsicness Premise outright, on the grounds that what prudentially matters is, like identity, just one of the many relations that is extrinsic. In "Reasons and Reductionism," Mark Johnston considers this strategy, though he ultimately rejects it. Johnston writes:

Whence the plausibility of the crucial principle that whether one has reason to be specially and directly concerned about some future person depends only on the intrinsic aspects of the relation between oneself and that future person? Certainly not from the plausibility of the general claim that extrinsic features do not matter. We often take extrinsic features to be highly relevant to how we evaluate some fact or relation (Johnston 1992, 609).

Johnston goes on to cite the examples of "exclusive ownership, winning, unique achievement, and intimacy" (cf. Sosa 1990, 319–320). With the possible exception of intimacy, all of these are clearly examples of the sort of extrinsic relation Parfit takes to be paradigmatic. Exclusive ownership obtains when I bear an intrinsic relation of ownership to some object, and no one else bears the same relation; winning obtains when I reach the finish line and no one else has yet done so; and unique achievement obtains when I alone succeed in doing thus-and-such.

But it is hard to see how these analogies could convince anyone who does not already accept that rational prudential concern is a relation that can only hold one-one. Indeed, it is hard to see how they could be relevant, unless one were to stipulate that whether my relation to some future person gives me reason for prudential concern depends on the relation holding exclusively. Nothing about the concept of prudential concern seems to tell one way or the other²³. And nothing about the analogous examples seems to provide reasons for thinking that prudential concern is a relation with an explicit exclusiveness clause. So while appeals to these sorts of analogies may remind us that there are a range of relations which are extrinsically determined, as a means to establishing the crucial premise, they are question-begging at best.

4.4.2 An Unsuccessful Defense of the Intrinsicness Premise

But the simplest argument for the view that M is an intrinsic relation is equally question-begging. One reason one might think that fission preserves what matters for rational prudential concern is because being one of the two survivors of a fission transplant would feel, from the inside roughly like being someone who wakes up from a complicated operation. But whatever other reasons there are for accepting Parfit's analysis, the fact that a situation would seem, from the inside, qualitatively identical to ordinary survival does not give a reason to think that it is just as valuable.

²³ Cf. Sosa: "We can see the sorts of value that would be threatened by having too many spouses; but I for one have no inkling of what important values would be endangered by [division]...it is not easy to see what values would be endangered in such cases except only for the true survival of the mainstream protagonist" (Sosa 1990, 311).

This can be vividly illustrated by realizing that as far as perception “from the inside” goes, being the product of fission is no better or worse than being hooked up to what Nozick calls an experience machine (cf. Nozick 1974, 42-45), that is, a machine which would stimulate the brain and produce whatever set of experiences one would have in an ordinary life. What Nozick’s example plausibly shows is that something that appears from the inside to be worthwhile—such as seeming to undergo a certain sort of experience—might turn out upon reflection not to be nearly so valuable. So to the extent that intuitions converge in the fission case because readers imagine the case from the inside, these intuitions are not informative. Imagining the experience machine case from the inside would produce the same convergence of intuitions. If we would not trust them in the latter case, we should not trust them in the former.

A first retort to this objection might be the following. That a non-standard form of survival would feel from the inside like ordinary survival is at most a necessary condition of our valuing it. Against a background of veridicality, however, feeling the same from the inside becomes a much stronger indication of reasonable valuation. That is, if it feels the same from the inside, and it gives us accurate information about the world, then our desire to have this as a result of normal processes may be only derivative. Parfit seems to have something like this in mind when, in contending that it is irrational to prefer that continuation have its normal cause, he appeals to the following analogy:

Consider artificial eyes which would restore sight to those who have gone blind. Suppose that these eyes would give these people visual sensations just like those involved in normal sight, and *that these sensations would provide true beliefs about what can be seen*. This would surely be as good as normal sight. It would not be plausible to reject these eyes because they were not the normal cause of human sight (Parfit 1984/1987, 285, *italics added*).

That is, Parfit argues that we value our eyes only derivatively—as a source of visual sensation; that sight have its normal cause is not what matters to us. What does matter to us is that “these sensations would provide true beliefs about what can be seen.”

Similarly, what we care about in the case of continuation is not only that there will be someone whose inner life will feel the same to her as my inner life does to me, but also that her experiences will be, for the most part, veridical.

However, this response is not sufficient to draw a sharp line between the fission case and the experience machine case. Assuming that what we want when we require veridicality of experience includes memories of past actions as well as sensations of present ones, both forms of continuation have a non-veridical component. Just as every post-hookup experience of the person connected to the machine is false, so too, on Parfit’s view, is every I-involving memory of Lefty in the two-sided case²⁴. So Parfit’s analogy is inadequate. In the case of artificial eyes, both function and veridicality are preserved; in the case of fission, we have only the former. Against a background assumption of accuracy, phenomenological identity might justify equal valuation. But no such background conditions hold in the fission case. So the judgment that bearing a relation of prudential concern is rational for intrinsic reasons will be reliable only if it rests on more than imagining the case from the inside. It will have to account for the fact that veridicality of memory, normally valued for its own sake, does not matter in this case.

Of course, Parfit would retort that this is begging the question: the appropriate consideration, he would counter, is that Lefty’s quasi-memories be veridical, that is, that they accurately reflect someone’s actual experiences. Indeed, they don’t simply reflect someone’s experiences, they reflect the experiences of someone to whom he is related as a continuer, namely Brainy. Surely, Parfit would say, this is close enough to what we wanted when we

²⁴ This is so despite the fact that Lefty’s memories have their normal causes. Because Brainy has two continuers, neither of them is identical with Brainy (barring a multiple-occupancy view). So every (quasi-)memory that Lefty has that involves strict I-thoughts is, strictly speaking, false. One might try to get around this by replacing I-thoughts with I*-thoughts (see Rovane 1990), but this would be no help in terms of settling the question at hand. Precisely what is at issue is whether I*-thoughts are “just as good” as I-thoughts. (I thank Terry Irwin for showing me the need to include such a footnote.)

sought veridicality²⁵. But at this point we have reached a stalemate: quasi-memory “counts” only if identity doesn’t. And nothing about imagining the case from the inside can resolve that question one way or the other.

4.5 WHY IS THE FISSION ARGUMENT SO COMPELLING?

4.5.1 The Casewise Explanatory Difference Principle

The key to understanding the grip of Parfit’s argument is understanding why the Necessity and Intrinsicness principles seem so hard to reject—even once we have distinguished the Explanatory and Greatest Common Factor readings. What is it about M, and its relation to RPC, that inclines us to endorse these principles even when M is taken in the explanatory sense?

As the reader will recall, the principles in question are these:

The Necessity Principle: A’s prudential concern for B is rational (RPC) only if the relation which matters for rational prudential concern (M) holds between A and B.

The Intrinsicness Principle (for M): The relation that matters for rational prudential concern—M—is an intrinsic relation.

I think both principles gain their intuitive support from a principle slightly weaker than Necessity. Stated in terms of the relation between M and RPC, the principle is this²⁶:

The Casewise Explanatory Difference Principle (for M and RPC): If M is what explains RPC, then whether M obtains in a particular case will make a difference to whether RPC obtains in that particular case.

The Casewise Explanatory Difference Principle seems quite plausible. If one relation explains another, then it is reasonable to expect that the relations will co-vary. After all, if the second could

²⁵ I thank Nick Sturgeon for pressing me on this point.

²⁶ Of course, the principle can be stated more generally: If relation P is what explains relation Q, then whether P obtains in a particular case will make a difference to whether Q obtains in that particular case.

obtain without the first, or vice versa, then in what would the explanation of the second by the first consist? And the Casewise Explanatory Difference Principle shows us why the Necessity and Intrinsicness Principles seem compelling. It is a small step from Casewise Explanatory Difference to Sufficiency on the one hand (no M without RPC), and to Necessity on the other (no RPC without M); all that is required is the additional assumption that the difference made by the obtaining of M is dispositive. (Indeed, I accept the Sufficiency Principle in the explanatory sense largely for the reasons articulated by the Casewise Explanatory Difference Principle.) And from here, it is fairly easy to see what makes the Intrinsicness Principle appealing. If we grant, as I am happy to do:

(a) RPC obtains in the two-sided case

and if we accept, as the dispositive reading of the Casewise Explanatory Difference Principle tells us to:

(b) RPC cannot obtain without M

then it follows that:

(c) M obtains in both the one-sided and two-sided cases

so:

(d) whether M obtains between Brainy and Lefty is unaffected by whether M obtains between Brainy and Righty as well.

But it is hard to see what could justify our accession to (a) other than a commitment to the more general claim that:

(e) RPC obtains in any case which is intrinsically like the one-sided case

But from (e) and (b) follow:

(f) M obtains in any case which is intrinsically like the one-sided case

which captures as much of the Intrinsicness Principle as Parfit’s argument requires.

4.5.2 The Casewise Explanatory Principle and the Method of Agreement

The Casewise Explanatory Difference Principle is an application of a seemingly undeniable principle of scientific reasoning, what Mill called the Method of Agreement. Mill's principle says that: "If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon²⁷." The counter-intuitiveness of denying this principle is brought out by the following example. Suppose that whenever I strike a match against the side of a matchbox and say "let there be light," the match bursts into flame; whenever I strike a match against the side of a matchbox and say nothing, the match bursts into flame; and whenever I simply hold the match in the air and say "let there be light," the match remains unlit. It would be absurd to conclude from this that in the first case—when I strike the match and utter the phrase—that it is what I have said that has caused the match to burst into flame. If, whenever we have Y alone Z occurs, but whenever we have X alone Z does not occur, then we seem rationally compelled to conclude that when we have X and Y together and Z occurs, it is Y rather than X that does the explanatory grounding. Indeed, even in the absence of any evidence about what happens when X occurs without Y (that is, even in the absence of the incantation-alone case), if both Y alone and Y-plus-X consistently give us Z, we seem compelled to conclude that it is Y (rather than X or the combination Y-plus-X) that explains the occurrence of Z.

4.5.3 Fission and the Method of Agreement

Let us follow Parfit in using the expression Relation R to refer to psychological continuity with the right sort of cause (the "right sort" with regard to supporting rational prudential concern, that is)²⁸. With this in place, we can easily articulate the connection

²⁷ Mill I:451. Strictly speaking, the principle which concerns me is a cousin to this principle; I am concerned with justification rather than cause and effect.

²⁸ Cf. Parfit 1984/1987, 215. I neglect the distinction Parfit makes between psychological continuity and psychological connectedness (cf. Parfit 1984/1987, 206), since the details are unnecessary to my discussion above.

between the Method of Agreement and the case with which we are concerned.

Grant, for the sake of argument, that personal identity over time can be analyzed as the non-branching holding of relation R: Y is identical to X iff Y bears Relation R to X, and no one else bears Relation R to X. I conceded in section 2 that whenever both Relation R and identity obtain, rational prudential concern is warranted. And I also conceded there that in the two-sided case—where Relation R obtains without identity—rational prudential concern may be warranted. And given the analysis of identity that I have just granted, I have also conceded that rational prudential concern is never warranted in the absence of R. So if we apply Mill's method of agreement to this case, it seems clear that even when both identity and Relation R obtain, it is Relation R that matters for prudential concern²⁹. How, then, can I possibly claim that when both identity and Relation R obtain, it is identity that matters (in the explanatory sense)³⁰?

What I will try to show in the next section is that we can do precisely this: the Method of Agreement is false when applied to certain cases of explanation, and hence the Casewise Explanatory Difference Principle and the Intrinsicness Principle (for M) are false.

²⁹ Putting things this way is a bit misleading. The way in which identity differs from Relation R is unlike the way that saying "let there be light" is something different from striking the match against the side of the box. The fact that a relation of personal identity holds just is the fact that Relation R holds in a non-branching fashion. But this asymmetry is irrelevant to the discussion at hand.

³⁰ Cf. Parfit: "In all ordinary cases, personal identity and [R] coincide. When they diverge, [R] is what matters. That strongly suggests that, in all cases, [R] is what matters...If, when two facts come apart, one of them is what matters, why think the other is what matters when they coincide?" (Parfit 1994, 38).

4.6 HOW ABSENT FEATURES CAN BE EXPLANATORY

4.6.1 Human Bodies and Borrowed Luster³¹

Let's begin with the following example. To the extent that respect is an attitude constrained by rationality, it is surely rational to treat a living human body with respect; even in cases where our actions will not cause pain to another, we bear certain obligations towards living human bodies that preclude our treating them with wanton disregard. It is also rational, to the extent that respect is an attitude constrained by rationality, to treat a human body—even if it is non-living—with respect; again, we bear certain obligations that preclude our treating them with wanton disregard. But the living human body has a feature that the non-living body lacks, namely, being alive. So the Method of Agreement suggests the following analysis of this case: Since whenever there is something that is a human body that is living ($X + Y$), it is rational to treat that thing with respect (Z), and whenever there is something that is a (mere) human body (X), it is also rational to treat that thing with respect (Z), then what explains the rationality of respect in both cases (X and $X + Y$) must be the simple fact that the entity is a human body (X).

It seems to me clear that this analysis is false: what makes it rational to treat a non-living human body with respect is that it is rational to treat a living human body with respect, and not, as the

³¹ I have found that this example is extraordinarily persuasive to some readers, and utterly unconvincing to others. If you are one of the latter, bear with me. The phenomenon I am describing is sufficiently widespread that little turns on the force of this particular case.

Method of Agreement suggests, the other way around³². But this example is not unique; the Method of Agreement gives the wrong answer in a whole class of cases. Whenever we are concerned with the explanation of attitudes towards states of affairs imbued with symbolic utility³³, the Method of Agreement gets the order of explanation backwards: the proper treatment of stuffed animals may be understood as a function of the proper treatment of real ones; engaging in a certain ritual action may be significant because of how it fits into a larger pattern of practices; and not eating a particular sort of food on a particular occasion may be imbued with gravity as part of a more global dietary commitment. Rational prudential concern in cases of bare R-relatedness is, I contend, best understood in the same way: what makes prudential concern rational in the double-transfer case is the fact that in ordinary cases, R-relatedness is sufficient for identity.

4.6.2 Explaining Valuation

Again, let us begin with an analogy. Suppose we are presented with a recording that sounds exactly like a recording of some great cellist playing Bach. And suppose further that in the actual world, there is only one way such a recording could have come into existence: by actually being a recording of some great cellist playing

³² One might object to this example on the grounds that what deserves respect is not a non-living human body, but rather a dead body—one that once embodied the life of some particular person. And having-been-the-embodiment-of-a-particular-person is not a feature that is absent from the non-living body; rather, it is a property that the body has in virtue of its causal history*. (I thank Terry Irwin for this objection.) My reply is two-fold. First, for reasons that I discuss below, I think an attitude of respect might well be rational even towards a non-living human body that was never alive, should one come into being by some non-natural process. In such a case, the Method of Agreement would certainly get the order of explanation the wrong way around. But even without this concession, my point still goes through. If we suppose that having-ever-been-a-living-human-body applies to all human bodies, it is still true that being living is what, at base, explains the rationality of our valuing bodies as such. So again the explanatory work is done by an absent feature.

*One might also wonder about the case of a not-yet-living body, such as a fetus. My answer here is parallel, *mutatis mutandis*.

³³ For a discussion of symbolic utility, see Nozick 1993, especially 26–35.

Bach³⁴. Nonetheless, we would be “making the mistake of the untraveled” to suppose that this is the only way things could be. Surely it is possible that a bunch of computer sound technicians playing around in a CD studio could produce a recording with identical sound-patterns. Let’s call this latter case the sound-alike and the former case the actual recording. What I want to argue in the next two sections is the following:

- (a) it might well be rational to value the sound-alike just as much as we value the actual recording,

and

- (b) what makes (a) true is that it is rational to value the actual recording as we do.

That is, even if we value the sound-alike recording as much as we value the actual recording, we need not conclude—as the Method of Agreement would require—that what explains our valuation is whatever feature is common to both cases.

So why might it be rational to give full value to the sound-alike recording even if it is not actual³⁵? The answer, presumably, is that it would provide us with precisely the same sort of aesthetic pleasure that the actual recording does; experientially, hearing the

³⁴ The case I am imagining differs somewhat from the one described by Nelson Goodman in “The Perfect Fake” (Goodman 1976, 99-102). Goodman supposes that “we have before us, on the left, Rembrandt’s original painting *Lucretia* and, on the right, a superlative imitation of it. We know from a fully documented history that the painting on the left is the original; and we know from X-ray photographs and microscope examination and chemical analysis that the painting on the right is a recent fake” (Goodman 1976, 99-100). Goodman is concerned with the question of whether there could be “any aesthetic difference between the two pictures for x at t , where t is a suitable period of time, if x cannot tell them apart merely by looking at them at t ” (Goodman 1976, 102). In my example, there is—by stipulation—no qualitative difference between the authentic recording and the sound-alike; the only difference between them is their causal history. Nonetheless, aspects of Goodman’s discussion are relevant to my point. See especially Goodman 1976, 105-106.

³⁵ A puzzle: Nelson Goodman distinguishes between works of art that are autographic and works of art that are allographic. Autographic works are works for which “even the most exact duplication of it does not thereby count as genuine;” allographic works are works which are not autographic. Goodman contends that painting is autographic (even a perfect replica of a Rembrandt is not a Rembrandt), whereas music is allographic (even a beginning pianist’s performance of a Chopin Etude is a performance of Chopin). Goodman concedes that “there may indeed be

Continued on next page

sound-alike would be exactly the same as hearing an actual recording. By stipulation, there is no qualitative difference between the two; no matter how perfect one’s musical sense, there would be no discernible discrepancy between them. Given the sort of thing aesthetic value is, what else is there to say?

Despite its appeal, this sort of reason for accepting (a) is surely too strong. It suggests that all that matters in judgments of value are our experiences, that underlying causes are irrelevant to such assessments. But there are at least two reasons that might seem mistaken to judge the value of an entity without attending to the process by which it came to exist. The first is that we might care about the process itself for process-dependent reasons, thinking it valuable that someone perform the actions that bring the entity into being via the normal route. (On these grounds, we might prefer homemade brownies to equally good brownies from a bakery, or a union-made shirt to an identical one produced in a non-unionized factory.) Secondly, we might care about the process because we think it valuable that we respond to the world in a way that tracks what is actually going on. So we might care about how an entity came into being for epistemic reasons, because we want to know that we can take certain things as evidence. (On these grounds, we might be disturbed by apparent photographs produced by means of Digitized Image Technology, or by particularly effective forgeries.)

Both of these worries are at play in the music case. We think that endeavors like practicing the cello and composing music and crafting high-quality instruments are valuable endeavors in themselves. We also want to know that we are justified in taking a great recording as evidence for the fact that a great musician produced it, and we want our evaluative standards to reflect this desire.

forgeries of performances...that purport to be by a certain musician etc.; but these, if in accordance with the score, are nevertheless genuine instances of the work” (Goodman 1976, 113; cf. 118). In my example, however, even though the sound alike is “in accordance with the score,” I think it is not a “genuine instance of the work.” Nor is it properly speaking a forgery. “A forgery of a work of art is an object falsely purporting to have the history of production required for the (or an) original of the work” (122). But there is no sense in which the sound-alike that I describe purports to have one history or another. Perhaps the difficulty we have in classifying this case can be traced to the fact that it violates our ordinary assump-

It seems difficult to see how we could respect these intuitions while endorsing (a). But if we accept (b), we can have it all: we do justice both to the intuition that leads us to say that the sound-alike is as valuable as the actual recording, and to the intuitions that lead us to say that it is not.

4.6.3 Exceptions, Norms and Local Adaptation

Imagine that the exceptional case is the norm, and that things that sound like recordings of great cellists playing Bach are not, in general, recordings of great cellists playing Bach. If you accept that under such circumstances we might value such musical experiences less—or at least differently—than we do now, then I have a foothold for my final argument.

Here is the argument. If we say that our valuing the common feature (what the recording sounds like) is what explains our valuing the enhanced feature (the recording actually being of such-and-such), then the only way we can account for such a change in value would be on the grounds that there would likely be less music-producing activity. But this doesn't seem to capture the extent of the transformation. Surely it is not only because there would be less cello-practicing that our attitudes would change. On the other hand, if we say that the enhanced feature—being an actual recording—is what explains our valuation of the common feature, then we can account for why it is that in the face of such global disruptions, we would be inclined to rethink our valuational scheme. We can account for the fact that in a world where all music was produced in that way, it might be rational to value music less than we do, even if in the case where such a process is exceptional, we would not discount a recording simply because it had an abnormal cause. The reason is that the case we have been considering is one in which the interpretation of the exceptional case is parasitic on the interpretation of the normal cases from which it deviates. To the extent that we are able to account for the exceptional case, our accounting takes the form of a sort of local adaptation; we maintain our background assumptions about recordings in general, and adapt our standard responses to the case at hand.

There is reason to think this phenomenon of local adaptation is quite general. Consider, for example, what has happened to the concept of motherhood in the face of recent technological advances. As it has become possible to implant the egg of one

woman into the uterus of another, a previously unnoticed distinction has been drawn between genetic mother on the one hand, and birth mother on the other³⁶. Since the cases are exceptional, there has been an effort to “save” the concept; and since it is the egg-donor whose genetic information is carried on to the child, it is the birth mother who has been given the status of “surrogate.” But were the practice to become widespread, the concept of motherhood would break down entirely. We would no longer have the idea of filial concern for one's mother as such, because there would be no unitary concept of “mother” that lay behind it. We might well have two similar concepts—filial concern for one's birth mother and filial concern for one's genetic mother; and we might well think these concepts were more similar to one another than either one would be to the concept of concern for one's child, or one's spouse, or one's sibling. But I think that in such circumstances, the concept of maternal-filial concern simpliciter would have no application.

4.6.4 Prudential Concern in a World of Fission

Here, then, is the final piece of the puzzle. In a world in which fission was the norm, I contend, there would not be a concept of pru-

³⁶ Consider the following parody of a new-age greeting card, taken from *The New York Times Sunday Magazine*:

You're Someone Special
Mom...
Gee, it feels funny to call you that.
But after all, you are the woman
who brought me to term.
And even though it was “just a job,”
I feel as though we have a lasting bond.

I know it can't have been easy
carrying around someone else's baby,
especially a big eater like me!
So I just want to say,
thanks for being my birth mother!
The time we spent together
will always mean something special to me.

(Rubiner 1996, 60)

dential concern of the sort Parfit needs for his argument to succeed. There might be a somewhat similar concept, such as the concept of concern-for-one's-R-related-continuer. But there is reason to think that it would not be the same concept as the one we have, the concept that describes the relation we bear to our future selves³⁷. In such a world, Brainy's relation to Lefty in both the single-transfer and double-transfer cases would indeed contain what matters for concern-for-one's-R-related-continuer. But the relations would not contain what matters for prudential concern, because there wouldn't be prudential concern in the relevant sense³⁸.

Now, if we say, as I have been arguing, that identity is what matters in the explanatory sense, then we have some way to account for the fact that our concepts might well change in the face of such a global disruption³⁹. But the same option is not available if we say that what matters is R-relatedness. If it is (mere) R-relatedness that explains our valuation of identity, rather than the other way around, then the global replacement of identity by mere R-

³⁷ But why is this not like the story of the Oxford misogynist, who, at a debate over whether All Souls College should admit women, contended that All Souls College could not admit women, for if it were to do so, it would cease to be All Souls College? (I thank Derek Parfit for the story, and for the objection it raises.)

The answer, I think, is that the role played by All Souls College before and after the admission of women as fellows would be largely unchanged, whereas the role played by R-continuer concern in such a world would be too different from the role played by prudential concern in our world for the two to be considered the same concept. To argue fully for this would require another paper (at least), but the line of thought is as follows. The concept of prudential concern is tied up with concepts of fairness, responsibility, justice, and rationality. Our views about the sorts of rational and moral obligations we have to ourselves and others considered as beings who exist through time rest on the assumption that each of us will have at most one continuer, and that that continuuer is someone with whom we will be identical. Disruption of this background assumption would result in disruption of the entire framework by which we make sense of this wide range of concepts. And to the extent that prudential concern is interconnected with them, it too would be disrupted. (For discussion of some of these issues, see Diamond 1991, Korsgaard 1989; Rovane 1990 and 1998; Schechtman 1996; Whiting 1986; and Wolf 1986.)

³⁸ Something like this seems to be what McDowell is suggesting when he writes: "According to the view I recommend, a context of facts about the objective continuation of lives helps to make intelligible a face-value construal of what Locke actually says, that continuous 'consciousness' presents an identity through time" (McDowell 1997, 234).

³⁹ Note that even if only this weaker claim is true—that there might not be a concept of prudential concern—my argument can still go through.

relatedness should make no difference to the value we place on the relation we bear to our continuers⁴⁰. If we follow Parfit in accepting that what matters for prudential concern in the explanatory sense is not identity, we have no way to account for the fact that prudential concern as we know it might not exist under the conditions I described. If, however, we maintain that what matters for prudential concern is identity, then we are able to account for such a potential disruption. By properly recognizing the way in which contingent features play a role in the organization of our concepts, and the way in which exceptions of this sort depend on a background of normal cases, we are able to account for the case with appropriate provincialism, and not to "fall into the extravagances of the knights of romances" (Descartes, I.8).

⁴⁰ It may seem that I have only succeeded in showing something rather weaker, namely that what matters is R-relatedness plus uniqueness-at-a-time, and not identity as such. (Sydney Shoemaker has raised this objection in conversation; a similar case is described in Raymond Martin's "Fission Rejuvenation.") Imagine a world in which everyone undergoes fission at the age of 20, but where one of the resulting survivors always dies within a few days (Shoemaker's case), or where one of the survivors is put into a comatose state, and "awakened" only upon the death of the other (Martin's case). It seems plausible to say of such a world: (a) that there would be prudential concern, and (b) that as a rule, this prudential concern would not have to be for someone with whom one was identical. This seems to show that what matters is being the currently-unique long-term survivor (and not being identical).

It seems to me that the correct sort of reply to these cases comes in two parts, which together show that the more compelled one feels to accept (a), the less compelled one thereby feels to accept (b) and vice versa. (This is only the sketch of an argument; it omits many details.) Case one: if the example supposes that the quick death or coordinated coma of one of the fission products is guaranteed by certain sorts of laws that govern the imagined world, it seems plausible to say that pre-'fission' persons are identical to their post-'fission' survivors. Their way of surviving over time differs from what we would consider normal, but the logic of identity continues to govern the relations in question, and so (b) can be denied. Case two: If, on the other hand, there is the always-present but never-realized live possibility of multiple-continuation, then I think the concept of prudential concern would begin to break down, for the reasons adverted to in footnote 37. (Note also that, even in the worlds hypothesized, identity is still the norm: until the time of fission, B is (strictly) identical to her future stages, and after fission, both L and R are (strictly) identical to theirs.)

4.7 CONCLUSION

4.7.1 Summary

I began this chapter by suggesting that there is a danger to philosophical inquiry that ignores what I have been calling the facts of life. That human beings come into existence only through the predictable sequence of events that I described in section 1 is one of the background truths against which we organize our concepts. At the same time, it seems possible that there could be circumstances—fission is one—where a process that is ordinarily identity-preserving would turn out to be entity-creating. That is, it seems possible that there could be a process with the following character: if it happened in one way (what we have been calling the “single-transfer case”) it would result in the continued existence of some entity over time; but if it happened in another way (what we have been calling the “double-transfer case”) it would result in the creation of two new entities⁴¹. But if the entities in question are self-conscious, as human beings are, then this possibility raises the following puzzle. To the extent that the process itself would—intrinsically—be the same in both cases, how could the rationality of one’s attitude towards one’s continuer depend on whether the process ended up being identity-preserving, or whether it ended up producing two new human beings? Presumably one’s attitude towards one’s continuer would—rationally—be the same in both the single-transfer and the double-transfer case. And with this much, I said I agree.

The question that has concerned me in this chapter has been the question of what lessons can be drawn from this fact. Parfit contends that from it, we can conclude that what makes my prudential concern for myself tomorrow rational is not the fact that myself-tomorrow will (presumably) be identical to myself-today, but only that she will be connected to me by the right sort of causal process that will result in the right sort of relation of psychological continuity and connectedness. I have tried to show that this

⁴¹ Of course, part of what is at issue is whether it is correct to describe the cases as involving the “same process.” But I trust that despite the sloppiness of my language, my meaning is clear.

conclusion can be blocked. I pointed out that Parfit’s reasoning rests on two principles: the Necessity Principle, and the Intrinsicness Principle (for M), and that the intuitive force of these principles can be traced to a tendency to endorse what I called the Casewise Explanatory Difference Principle, which is a special case of what Mill called the Method of Agreement. But while the Method of Agreement is undeniable when we are concerned with scientific reasoning, it cannot be straightforwardly applied to cases of explaining value. In such cases, the Casewise Explanatory Difference Principle is false, as are the Necessity Principle, and the Intrinsicness Principle (for M).

What this means is that Parfit’s argument shows much less than he takes it to show. It shows only that there are conceivable circumstances where it might be rational to bear a relation of prudential concern towards a continuer with whom one was not identical. But it does not show that identity is not what—in the explanatory sense—matters.

4.7.2 Larger Lessons

Although most of my argument has focused on a single example, I should make it clear that I take my discussion to have general implications. In the case I described, we are asked to consider a scenario in which a pair of features that coincide in all actual situations are imaginatively separated, and to make a judgment about which of the two features has conceptual primacy. I have argued that the proper interpretation of the case is precisely the opposite of what it has generally been taken to be. And I think the reason its implications have been so misunderstood is this: certain patterns of features which coincide only fortuitously may nonetheless play a central role in the organization of our concepts. To the extent that imaginary scenarios involve disruptions of these patterns, our first-order judgments about them are often distorted or even inverted.

CHAPTER 5

Conclusion

At the beginning of this study, I raised the question of how thinking about imaginary cases can help us to learn new things about the world. I went on to suggest that the answer has something to do with the role played by exceptional cases. I then identified a structure which I suggested is common to thought experiments in science and philosophy. And I closed the introductory chapter by submitting that the question has at least three different facets, one concerning factive, one concerning conceptual, and one concerning valuational thought experiments.

In the next three chapters, I presented detailed case studies of three thought experiments, each representing one of the three sorts of case. And in each of the case studies, I tried to establish a number of specific claims. In the first case study, I discussed a famous thought experiment of Galileo's, and tried to show that the guided contemplation of an imaginary scenario can provide us with new scientific knowledge in a way that argument alone cannot. In the second case study, I tried to show that standard interpretations of the puzzle of the Ship of Theseus founder because they ignore the importance of the background norms against which we can make sense of local instances of extrinsically-determined identity. And in the third case study, I tried to show that certain thought experiments in the personal identity literature are inconclusive because they disregard the explanatory role played by contingent facts about the ways human beings come into existence.

The aim of this conclusion is to tie these strands together. I will do so by discussing each of the thought experiments in turn. What

I try to show is that exceptionality plays a role in their evaluation at three distinct levels: (a) in the patterns of classification of certain states of affairs as normal or exceptional; (b) in the ways in which the particular exceptional (that is, unusual) situation described by the scenario should be accounted for; and (c) in the ways in which this exceptionality is accounted for at the level of explanation. It is the interplay among these three aspects of exceptionality that explains the respective successes and failures of the particular cases I have discussed.

5.1 FACTIVE THOUGHT EXPERIMENTS: GALILEO

Scientific thought experiments are typically factive¹; they are attempts to elicit physical intuitions about what would happen under certain conditions. Such thought experiments are puzzling because they seem to describe cases where we learn something new about the physical world, even though we have no new *empirical* information about the world².

Three sorts of responses have been offered to this puzzle. In Chapter 2, I discussed the *empiricist* response, as exemplified by the writings of John Norton (Norton 1991, 1996). Norton contends that scientific thought experiments have demonstrative force only insofar as they are arguments in disguise. Otherwise, he contends, we would not be able to account for the fact that they provide us with new information that is not new *empirical* information. In Chapter 2, I argued that such a view cannot account for Galileo's famous thought experiment, where as a consequence of thinking about a particular imaginary case, the Aristotelian is brought to reject certain framework-defining assumptions about the sort of property natural speed is. Prior to thinking about the case that Galileo describes, the Aristotelian is committed to the view that natural speed is a function of weight. After thinking about the case in the way that Galileo encourages, he recognizes that such a conception is untenable. He comes to

¹The reader will recall that *factive* thought experiments are cases where an imaginary scenario is described, and the reader is asked to think about *what would happen* in such a case.

² Thomas Kuhn poses the problem concisely as follows: "How...relying exclusively upon familiar data, can a thought experiment lead to new knowledge or to a new understanding of nature?" (Kuhn 1977, 241). (Cf. also the discussion of Norton in Chapter 2 above.)

understand that other things he knows about the ways that middle-sized physical objects behave belie his previous representation of natural speed.

This realization is not a sudden flash of Platonic inspiration, which illuminates previously invisible *a priori* physical truths. Advocacy of such a view can be found in the writings of James Robert Brown, against whom Norton directs his argument. Brown contends that in some cases, "thought experiments give us (fallible) *a priori* beliefs of how the physical world works. With the mind's eye, we can see the laws of nature" (Brown 1991, 155)³. But where there is no need to posit such mysterious entities, they ought to be avoided. And, as I have argued above and will argue below, there is a satisfactory explanation of the phenomenon which does not make appeal to such enigmatic abstracta.

Between immoderate empiricism and immoderate platonism lies a third sort of explanation. Eschewing the assumption that there is sharp line between theory on the one hand, and raw data on the other, the position introduces an element of *constructivism*. It suggests that the new knowledge in the Galileo case comes neither from argument nor from insight, but from a reconfiguration of conceptual commitments on the part of the Aristotelian which enables him to see old phenomena in a new way⁴.

³ Brown considers the thought experiment of Galileo's which we have been discussing to be one of the two "best examples of...a platonic thought experiment" (Brown 1991b, 77); the other is the 1935 Einstein, Podolsky and Rosen thought experiment or "EPR".

⁴ The three proposals I have been considering are *explanatory* rather than *genetic*. Each would be compatible with the proposal made by Roger Shepard in his William James lectures. Shepard suggests that "every thought experiment is preceded by an enormous number of actual experiments," namely, the selectively-relevant experiences had by our ancestors in the long course of evolution of the species. (See Shepard 1994). So, he suggests, the knowledge gained as a result of thought experiment is knowledge of naïve physics implanted in us by evolution.

I am not sure Shepard's answer gives the whole story. But it is surely true that some knowledge of naïve physics is hard-wired. Both folk psychology and cognitive science tell us so; and there are additional arguments (some due to Shepard himself) which seem to refute convincingly the proposal that each of us is born a *tabula rasa*. But as with the discussion of Mach above (see Section 2.3.3), I have little more to add here; the question lies outside the scope of this study.

The thought experiment directs the Aristotelian's attention to inadequacies in his conceptual scheme, inadequacies which he recognizes to be inadequacies when he performs a particular experiment in thought. The experiment in thought involves asking himself what sorts of objects and motions there are in the world. Through this, he realizes that entification is not fixed by the world, whereas natural speed is. Such an explanation appeals to exceptionality at three levels, which I now discuss.

As I noted in the text, the Aristotelian needs to explain as aberrant what the Galilean takes as typical: that the rate of fall of bodies of radically different weights is sometimes (nearly) identical (see section 2.3.1). And as I went on to note, he does so by appealing to additional principles which show how the world's complexity has prevented the true principles from being manifest; he suggests, for instance, that from great enough heights, the relevant differences would be apparent. Part of the power of Galileo's thought experiment comes in helping the Aristotelian to see that the cases he has been taking as anomalous are in fact the norm, that what cries out for explanation is precisely the opposite of what he initially thought; the cases which require explanation are those where heavier objects fall faster than lighter ones. So the first way in which exceptionality comes into the explanation is that the Galilean thought experiment brings the Aristotelian to see that the cases he has been taking to be exceptional are in fact typical, and *vice versa*. We'll call this *exceptionality in the first sense*.

Secondly, the specific case which Galileo asks the Aristotelian to contemplate is exceptional in the following sense. In ordinary cases involving falling bodies, the objects which fall are objects for which criteria of individuation are clear. We ask: what would happen if we were to drop thus-and-such body from this-and-that height? Or: what would happen if we took these two things up to the top of that tower and let them go? But the case which Galileo asks the Aristotelian to think about is one where the criteria of entification are not so obvious. Is that strapped-together thing one body or two? The world doesn't really tell us.

But despite the fact that the object in the example is odd in this way, the conclusions which can be drawn from thinking about what would happen to it are general conclusions. Descriptive physical science aims at generality, and if the theory must be modified to account for a case which is (in this second sense) exceptional,

then the theory must be modified in general. That it is here appropriate to allow exceptions to drive norms is confirmed by the second test (see Section 1.1.3). Even if all objects were like this strapped-together object, it would still make sense to have a theory about the natural speed of falling bodies⁵. So the second way in which exceptionality comes into play is in bringing out why the evaluation of the imaginary scenario is rightly taken to reveal general truths about cases beyond the scenario (see Section 1.3.2).

The third and most interesting way that exceptionality comes in is at the level of the explanation itself. In Chapter 4, I argued that absent features play explanatory roles by providing the background assumptions against which explanations make sense; that is, when we are concerned with explanations, we are concerned with norm-driven exceptions. What this means in the Galileo case is that what I above called *exceptionality in the first sense* (that the inversion of exceptional and normal in the order of explanation) is sufficient to bring with it new knowledge. Let me explain.

Galileo's thought experiment brings the Aristotelian to see that the cases he had been taking as remarkable are in fact ordinary, that what demands explanation is precisely the inverse of what he previously supposed. In one sense, this transposition is trivial; all that has happened is that two classes of cases have been differently diagnosed. But if we take seriously what we have learned about exceptionality in the third sense, then it becomes apparent that this inversion has far-reaching consequences.

As I argued in Chapter 4, exceptions in explanation are norm-driven; we can make sense of anomalous cases only against a background of regularity, and we recognize these as cases of the relevant sort only, so to speak, out of courtesy. The Galilean thought experiment helps the Aristotelian to see that explanation is needed for cases where two bodies of differing weights fall with *different* natural speeds (and not cases where two bodies of differing weights fall with the same natural speed). And what this means is that it helps him to see that the cases where bodies fall with different natural speeds need to be understood *through the lens* of

⁵ Indeed, in certain presentations of modern physics, all bodies *are* conceived of as strapped-together in this sense—that is, as particles held together by adhesive forces. (I thank Simon Saunders for pointing this out to me.)

cases where bodies fall with the same natural speed. They need to be seen as *exceptional*; more specifically, they need to be seen as exceptional with respect to a norm that says: as a rule, bodies fall at the same speed regardless of their weight.

And this is the deepest sense in which the contemplation of an imaginary scenario forces the contemplator to make sense of an exceptional case, and thereby brings him to new knowledge. Because of the relation of exceptions and norms in the case of explanation itself (third sense), because of the relation of exceptions to norms in the case of physical theories (second sense), and because of the inversion of exceptions and norms in what demands explanation in this particular case (first sense), it can truly be said of this thought experiment that it brings the Aristotelian to see new things in the world; it does so by bringing him to see the world in a new way, against a newly configured background of norm and exception.

5.2 CONCEPTUAL THOUGHT EXPERIMENTS: THE SHIP OF THESEUS

Conceptual thought experiments are cases where an imaginary scenario is presented, and the reader is asked to think about *how such a situation should be described*. In Chapter 3, I discussed a particular conceptual thought experiment, that of the Ship of Theseus; I argued that the implications of that thought experiment have been standardly misunderstood.

We might express the problem using the vocabulary of Section 1.3.2. Clearly, the scenario described is imaginable; it is an example of what I called in Section 1.2 a *hypothetical* case, which means that for all we know, such a case might actually have occurred. So there is no reason to level an *unimaginability* criticism against the case. Moreover, standard analyses of the scenario, to the extent that they recognize that the case involves pitting two principles of identity against one another, are largely reasonable. For the most part, there is no reason to level *unsound argument* criticisms against interpretations of the case⁶. Instead, the problem with dis-

⁶ Dissolutions like those offered by Parfit and van Inwagen suggest that apparent depth of the puzzle is a result of false assumptions about the relation of our concepts to the world. Such positions err not in their reasoning about the particular case of the Ship of Theseus, but in the conclusions they draw about such cases in general. Solutions like those offered by Hirsch and Nozick give what I take to be the right answer to the question; where they err is in the generalizations that they make on its basis.

cussions of the Ship of Theseus is in the *application* of the lessons of this particular case to cases of identity in general. The appropriate level at which to object to the solutions presented in Sections 3.5 and 3.6 above is on *inapplicability* grounds.

Returning to the vocabulary of this conclusion, we might say that the Ship of Theseus is an exceptional case in the second, flat-footed sense; that is to say, it is an unusual case. What should we make of this? In the Galileo example, the unusual case of the falling strapped-together body served to establish that the speed at which such bodies fall is independent of their weight. And the proper conclusion to draw on the basis of the case was that the previous classification of normal and exceptional cases was inverted. Heavy and light bodies falling at the same speed are normal; heavy and light bodies falling at different speeds are anomalous. Because the proper way to account for the exceptional case is by changing the norms to meet the exception, the proper way to understand the patterns of exceptionality is to reverse the previous assumptions.

In the story of the Ship of Theseus, the unusual case presents us with an instance where a process which is ordinarily entity-preserving is instead entity-creating. But here it is not true that the proper conclusion to draw is that such processes are in general entity-creating (or that they provide dubious mechanisms for identity-preservation). And the *reason* that this is the wrong conclusion to draw is again brought out by the generalization test (see Section 1.1.3). For if cases like Theseus were the norm rather than the exception, it would not make sense even to speak of identity-candidacy for ships. Ships would be like amoebae or cloud formations or World Wide Web sites—messy sorts of entities with obscure criteria of individuation and persistence; in general, we would not

We might classify the four views as follows:

| | Solution | Dissolution |
|-----------------------|----------------------|-------------|
| Closest continuer | Nozick Parfit | |
| Not closest continuer | Hirsch (traditional) | van Inwagen |

know what to say about their identity conditions over time. So the case of the Ship of Theseus is one in which the proper interpretation of the case is as a norm-driven exception. We can make sense of this anomaly, but only against a background of normal cases.

What, then, of the third level at which exceptionality enters the picture, the level of explanation? In the Galileo case, it was this level that solved our epistemological puzzle. It helped us to see why the perspective shift afforded by the accommodation of this exceptional case was sufficiently powerful to make pre-shift perceptions so radically different from post-shift perceptions that we might without exaggeration speak of *new* knowledge. In the case of the Ship of Theseus, where no such shift occurs, the consequences are more modest. I suggested that what the case reveals is that processes may be extrinsically determined in ways that we might not have realized. Explanation goes via normal cases; they provide the background against which we make judgments about the atypical. But the sort of radical shift in perspective that ensued as a result of the scientific thought experiment is not possible here. The fulcrum around which we seek to reorganize our conceptual commitments is not sufficiently independent of the commitments themselves to permit such a radical reconfiguration.

Thinking about the case in this way sheds new light on the errors of those who misgeneralize on its basis. The sorts of radical shifts in perspective which they propose are precisely the sorts of alterations we would expect if it were indeed reasonable to treat this as a case of exception-driven norms. So once again our schema has helped us to understand the import of an imaginary exceptional case.

5.3 VALUATIONAL THOUGHT EXPERIMENTS: PERSONAL IDENTITY

Valuational thought experiments are thought experiments where we are asked to make judgments about how we would *evaluate* what would happen in a particular imaginary scenario. In Chapter 4, I discussed a particular valuational thought experiment, the case of fission, and I argued that the case does not succeed in showing what Parfit takes it to show, namely, that personal identity is not what matters in justifying rational prudential concern.

In contrast to the other two examples, the case of fission is what I called in Section 1.2 a *counterfactual* case; even without looking

at the world, we know the case to be imaginary⁷. This means that the case must be approached with a certain degree of caution. Not only must we be careful not to misunderstand the implications of the exceptional case we are asked to consider, but we must be careful *in the interpretation of the case itself* not to import assumptions which would be inapplicable in that imaginary context. This suggests that the proper sort of criticism in the case of fission is at the level which in Section 1.3.2 I referred to as *unsound argument*. In making the judgment about the case which he does, Parfit offers an *evaluation* of the scenario which is fundamentally misguided. His conclusion rests on certain assumptions about how we would value certain states of affairs, assumptions which are illegitimate in light of the very scenario which they are meant to illuminate.

Let us examine this in light of the categories of the conclusion. Clearly, the case of fission is exceptional in the flat-footed second sense; division of the sort Parfit describes is meant to be extraordinary, not commonplace. And, as I argued extensively in the body of Chapter 4, the appropriate way to account for this case is to recognize that for it to make sense to us, it *must* be exceptional. Consider again the sound-alike recording. A recording that sounds just like a great cellist playing Bach is as valuable (as far as the sensory experience it provides) as an actual recording of a great cellist playing Bach. But the reason this is so is because, in all actual cases, such recordings *are* recordings of great cellists playing Bach. The value that attaches to the sound-alike is derivative; such disks shine like the moon, not like the sun. If the patterns whereby such recordings come into existence were disrupted, sound-alikes would

⁷ Baruch Brody considers this sufficient to make the case irrelevant. He writes “Parfit [and others]...see the very possibility of such a case as posing a conceptual crisis about identity. I say that only the actual occurrence of such cases would pose a crisis, and that it would be a scientific crisis about how to describe and explain what has occurred...[S]ince such cases do not occur, and it is only their occurrence that would cause a problem on my account, it follows that there is, at the moment, no splitting problem if my account is true” (Brody 1980, 59). For the reasons I discussed in the introduction, I do not believe we can be rid of imaginary cases so easily as Brody seems to think. But Brody is surely right to point out that there are *practical* questions raised by actual cases which are not raised by cases which are purely imaginary, and that changes in the ways that the world is *do* pose crises for concepts that we otherwise use unproblematically.

lose their value⁸. Our ability to judge their worth is parasitic on our ability to judge the worth of actual recordings.

What this means is that when we perform experiments in thought concerning the scenarios described in valuational thought experiments which involve counterfactual cases, our very judgments are filtered through the background patterns of fortuitous coincidence which play central roles in the organization of our conceptual scheme. Imaginary cases allow us to separate out apparently contingent concomitants. But some of our conceptual commitments may presuppose the coincidence of such concomitants, even though we can make sense of their coming apart.

When we make judgments about such cases, we do so standing atop the normal cases from which they deviate. But we cannot, as it were, throw away this ladder. We can see as far as we do standing where we do, but we should not mistake this for standing on the ground. Questions about how we would extend our practices are not ways of finding out what underpins our practices.

In the situations Parfit describes, we can make sense of the cases he adduces. But we cannot make sense of them *as* cases about persons. They tell us what sorts of beliefs and commitments we might expect to find among a group of individuals for whom different possibilities were live, but they do not tell us about our *own* beliefs and commitments. What seems to make these cases so puzzling is that the sorts of circumstances we imagine underpinning these live possibilities are circumstances which strike us as being at most technologically impossible⁹. But, as I have argued throughout, merely contingent constraints on the way things happen to be may play central roles in our categories of evaluation.

⁸ It might be that we would take them to have as much value as ordinary recordings now do. But if so, it would be for different reasons.

⁹ Consider the following analogy: Suppose there were a community in which it were possible to arrange for a total-matter-replacer to keep one's book collection from decaying. Each day, a quick vacuuming of one's bookshelves would result in all of the matter in the books being replaced by new matter, configured in an identical way, such that all of the observable features of the books—the color of the pages, the words on them, dents on the covers, penciled notes in the margins, post-it notes in the bibliographies—would remain the same. People faced with this option might well be rational to accept it, and it would seem at best unkind to say of them that they represented a culture that did not care about books. On the other hand, this does not show, I think, that all *we* care about when we care about some book is the qualitative features that the book manifests. Nor do I think it shows that all we care about when we care about some book is that it have the qualitative features it has brought about in the ordinary way because that's the only way we can get those features.

This brings us full circle to the Galileo case. There, we saw that the thought experiment is able to bring new knowledge to the Aristotelian by allowing him to see all cases involving falling bodies *through the lens* of cases where bodies fall with the same natural speed. This is what explains the thought experiment's power in the Galileo case. And in the fission case, precisely the same feature explains the thought experiment's limit. We cannot suddenly step outside the framework within which we ordinarily make sense of what we value and believe. The way the world is and the way we see the world are too deeply intertwined for this to be possible.

Bibliographies

6.1 BIBLIOGRAPHY OF WORKS CITED

Note: All references in footnotes and texts are to this Bibliography.

- Achinstein, Peter, and Owen Hannaway, eds. (1985). *Observation, Experiment, and Hypothesis in Modern Physical Science*. Cambridge, MA: MIT Press.
- Ackrill, J. L., editor (1987). *A New Aristotle Reader*. Princeton, NJ: Princeton University Press.
- Alston, William and Jonathan Bennett (1988). "Locke on People and Substances." *Philosophical Review*, Vol. 97 (1988): 25-46.
- Armstrong, D.M. (1983). *What is a Law of Nature?* Cambridge: Cambridge University Press.
- Aune, Bruce (1994). "Speaking of Selves." *Philosophical Quarterly*, Vol. 44, No. 176 (July 1994): 279-293.
- Baillie, James (1990). "Identity, Survival, and Sortal Concepts." *Philosophical Quarterly*, Vol. 40 (April 1990): 183-194.

- Baillie, James (1993a). "Recent Work on Personal Identity." *Philosophical Books*, Vol. 34, No. 4 (October 1993): 193-206.
- Baillie, James (1993b). "What Matters in Survival?" *Southern Journal of Philosophy*, Vol. 31, No. 3 (Fall 1993): 255-261.
- Baillie, James (1996). "Identity, Relation R, and What Matters." *Pacific Philosophical Quarterly*, Vol. 77, No. 4 (December 1996): 263-267.
- Bealer, George (1998). "The Autonomy of Philosophy." In DePaul and Ramsey (1998): 201-240.
- Beck, Simon (1992). "Should We Tolerate People Who Split?" *The Southern Journal of Philosophy*, Vol. 30, No. 1 (Spring 1992): 1-17.
- Beck, Simon (1998). "Back to the Self and the Future." *South African Journal of Philosophy*, Vol. 17, No. 3 (August 1998): 211-225.
- Bermúdez, José, Anthony Marcel and Naomi Eilan, eds. (1998). *The Body and the Self*. Cambridge, MA: MIT Press.
- Blakemore, Colin and Susan Greenfield, eds. (1987). *Mindwaves*. Oxford: Blackwell.
- Block, Ned, ed. (1981). *Imagery*. Cambridge, MA: MIT Press
- Bourgeois, Warren (1995). *Persons: What Philosophers Say About You*. Waterloo, ON: Wilfrid Laurier University Press.
- Brody, Baruch (1980). *Identity and Essence*. Princeton: Princeton University Press.

- Brown, James Robert (1986). "Thought Experiments Since the Scientific Revolution." *International Studies in the Philosophy of Science*, Vol. 1 (September 1986): 1-15.
- Brown, James Robert (1991a). *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. New York and London: Routledge.
- Brown, James Robert (1991b). "Thought Experiments: A Platonic Account." In Horowitz and Massey (1991): 119-128.
- Brown, James Robert (1993a). "Author's Response (to Norton [1993])." *Metascience* (new series), Vol. 3 (1993): 38-40.
- Brown, James Robert (1993b). "Why Empiricism Won't Work." *Proceedings of the Philosophy of Science Association* 2, 271-279.
- Brown, James Robert (1995). "Critical Notice of Roy Sorenson 'Thought Experiments'." *Canadian Journal of Philosophy*, Vol. 25, No. 1 (March 1995): 135-142.
- Bunzl, Martin (1996). "The Logic of Thought Experiments." *Synthese*, Vol. 106, No. 2 (Fall 1996): 227-240.
- Burke, Michael B (1980). "Cohabitation, Stuff and Intermittent Existence." *Mind*, Vol. 89 (July 1980): 391-405.
- Burke, Michael B. (1994). "Dion and Theon: An Essentialist Solution to an Ancient Puzzle." *Journal of Philosophy*, Vol. 91, No. 3 (March 1994): 129-139.
- Butler, Joseph (1736/1975). "Of Personal Identity." First Appendix to *The Analogy of Religion*. Reprinted with new pagination in Perry 1975a.
- Butts, Robert E. and Joseph C. Pitt, eds. (1978). *New Perspectives on Galileo*. Dordrecht/ Boston: Reidel.

- Care, Norman S. and Robert H. Grimm, eds. (1969). *Perception and Personal Identity: Proceedings*. Cleveland, OH: Case Western University Press.
- Carey, Susan (1985). *Conceptual Change in Childhood*. Cambridge: MIT Press.
- Carey, Susan (1993). "Cognitive Domains as Modes of Thought." In Olson and Torrance (1993).
- Carey, Susan and Elizabeth Spelke (1994). "Domain-Specific Knowledge and Conceptual Change." In Hirschfeld and Gelman 1994: 169-200.
- Cargile, James (1987). "Definitions and Counterexamples." *Philosophy*, Vol. 62: 179-193.
- Carrier, Martin (1993). Review of Horowitz and Massey 1991. *Erkenntnis*, Vol. 39 (1993): 413-419.
- Carritthers, Michael, Steven Collins and Steven Lukes, eds. (1985). *The Category of the Person: Anthropology, Philosophy, History*. Cambridge: Cambridge University Press.
- Cavell, Stanley (1979). *The Claim of Reason: Wittgenstein, Skepticism, Morality and Tragedy*. New York and Oxford: Clarendon Press.
- Chalmers, David (undated). University of Arizona. On-Line Personal Identity Bibliography.
<http://www.u.arizona.edu/~chalmers/biblio/3.html#3.7>
- Claggett, Marshall (1959). *The Science of Mechanics in the Middle Ages*. Madison: University of Wisconsin Press.

- Clavelin, Maurice (1974). *The Natural Philosophy of Galileo: Essays on the Origins and Formation of Classical Mechanics*. Trans. A. J. Pomerans. Cambridge: MIT Press.
- Clement, John (1983). "A Conceptual Model Discussed by Galileo and Used Intuitively by Physics Students." In Gentner and Stevens (1983): 325-340.
- Cooper, Lane (1935). *Aristotle, Galileo, and the Tower of Pisa*. Ithaca, NY: Cornell University Press.
- Damerow, Peter et al. (1992). *Exploring the Limits of Pre-Classical Mechanics: A Study of Conceptual Development in Early Modern Science: Free Fall and Compounded Motion in the Work of Descartes, Galileo and Beeckman*. New York: Springer Verlag.
- Dancy, Jonathan (1985a). *An Introduction to Contemporary Epistemology*. Oxford: Basil Blackwell.
- Dancy, Jonathan (1985b). "The Role of Imaginary Cases in Ethics." *Pacific Philosophical Quarterly*, Vol. 66 (January-April 1985): 141-153.
- Dancy, Jonathan (1993). *Moral Reasons*. Oxford: Basil Blackwell.
- Dancy, Jonathan (1997). *Reading Parfit*. Oxford: Basil Blackwell.
- Dauer, Francis W. (1972). "How Not to Reidentify the Parthenon." *Analysis*, Vol. 33, No. 2 (December 1972): 63-64.
- Davis, Philip J. and David Park, eds. (1987). *No Way: The Nature of The Impossible*. New York: W.H. Freeman and Company.

- DePaul, Michael R. and William Ramsey, eds. (1998). *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.
- Dijksterhuis, E. J. (1961, repr. 1986). *The Mechanization of the World Picture: Pythagoras to Newton*. Trans. C. Dikshoorn. Princeton, NJ: Princeton University Press.
- Donagan, Alan (1990). "Real Human Persons." *Logos (USA)*, Vol. 11 (1990): 1-16.
- Drake, Stillman (1977). "Galileo and the Career of Philosophy." *Journal of the History of Philosophy*, Vol. 38 (January-March 1977): 19-32.
- Drake, Stillman (1989). *History of Free Fall: Aristotle to Galileo*. Toronto: Wall & Thompson.
- Drake, Stillman (1990). *Galileo: Pioneer Scientist*. Toronto: University of Toronto Press.
- Drake, Stillman and I. E. Drabkin (1969). *Mechanics in Sixteenth Century Italy: Selections from Tartaglia, Benedetti, Guido Ubaldo, Galileo*. Madison: University of Wisconsin Press.
- Earman, John et al eds. (1993). *Philosophical Problems of the Internal and External Worlds: Essays on the Philosophy of Adolf Grünbaum*. Pittsburgh: University of Pittsburgh Press.
- Eilan, Naomi (1995). "Consciousness and the Self." In Bermúdez, Marcel and Eilan (1998): 337-357.
- Einstein, Albert (1961). *Relativity: The Special and General Theory*. New York: Crown Publishers.
- Fodor, Jerry A. (1971). "On Knowing What We Would Say." In Rosenberg and Travis (1971): 198-212.

- Fodor, Jerry A. (1998). *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford.
- Foot, Philippa (1978a). *Virtues and Vices*. Oxford: Blackwell.
- Foot, Philippa (1978b). "The Problem of Abortion and the Doctrine of Double Effect." In Foot (1978a).
- Gale, Richard (1991). "On Some Pernicious Thought Experiments." In Horowitz and Massey (1991): 297-304.
- Galilei, Galileo (1590/1960). *On Motion*. Madison, WI: University of Wisconsin Press, trans. I. E. Drabkin.
- Galilei, Galileo (1632/1967). *Dialogue Concerning the Two Chief World Systems*. Trans. Stillman Drake. Berkeley: University of California Press.
- Galilei, Galileo (1638/1914, repr. 1954). *Dialogues Concerning Two New Sciences*. Trans. Henry Crew and Alfonso de Salvio. New York: Dover.
- Galilei, Galileo (1638/1974, rev. 1989). *Two New Sciences, Including Centers of Gravity and Force of Gravity and Force of Percussion*. Trans. Stillman Drake. Toronto: Wall & Thompson.
- Gallagher, Shaun and Jonathan Shear, eds. (1999). *Models of the Self*. Exeter, UK: Imprint Academic.
- Gallagher, Shaun (undated). Canisius University, Buffalo. On-Line, Personal Identity Bibliography.
<http://gort.canisius.edu/gallagh/>
- Garrett, Brian (1985). "Noonan, 'Best Candidate' Theories and the Ship of Theseus." *Analysis*, Vol. 45 (October 1985): 212-215.

- Garrett, Brian (1988). "Identity and Extrinsicness." *Mind*, Vol. 97 (January 1988): 105-109.
- Garrett, Brian (1990). "Personal Identity and Extrinsicness." *Philosophical Studies*, Vol. 59, No. 2 (June 1990): 177-194.
- Gendler, Tamar Szabó (1998a). "Exceptional Persons: On the Limits of Imaginary Cases." *Journal of Consciousness Studies*, Vol., 5, Nos. 5-6 (1998): 592-610. Reprinted in Gallagher and Shear (1999): 592-610.
- Gendler, Tamar Szabó (1998b). "Galileo and the Indispensability of Scientific Thought Experiment." *British Journal for the Philosophy of Science*, Vol. 49, No. 3 (September 1998): 397-424.
- Gendler, Tamar Szabó (1999). "Review of Eric Olson's The Human Animal." *Philosophical Review*, Vol. 108, No. 1 (January 1999): 112-116.
- Gentner, Dedre and Albert L. Stevens, eds. (1983). *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gillett, Grant (1986). "Brain Bisection and Personal Identity." *Mind*, Vol. 95 (April 1986): 224-229.
- Goldman, Alvin (1989). "Psychology and Philosophical Analysis." *Proceedings of the Aristotelian Society*, 89 (1998-99): 195-209.
- Gooding, David C. (1992). "The Cognitive Turn, or, Why Do Thought Experiments Work?" In Giere (1992): 45-76.
- Gooding, David C. (1993). "What is Experimental about Thought Experiments?" *Proceedings of the Philosophy of Science Association*, Vol. 2 (1993): 280-290.

- Gooding, David C. (1994). "Imaginary Science." (Review of Horowitz and Massey (1991)). *British Journal for the Philosophy of Science*, Vol. 45, No. 4 (December 1994): 1029-1045.
- Goodman, Nelson (1976). *Languages of Art*, 2nd edition. Indianapolis: Hackett Publishers.
- Grethe, B., ed. (1996). *The Tanner Lectures on Human Values*. Salt Lake City, Utah: University of Utah Press.
- Grice, H. P. (1941/1975). "Personal Identity." *Mind*, Vol. 50 (October 1941). Reprinted with new pagination in Perry (1975a): 73-98.
- Hacking, Ian (1993). "Do Thought Experiments Have a Life of Their Own?" *Proceedings of the Philosophy of Science Association*, Vol. 2 (1993): 302-308.
- Häggqvist, Sören (1996). *Thought Experiments in Philosophy*. Stockholm: Almqvist & Wiksell International.
- Haksar, Vinit (1991). *Indivisible Selves and Moral Practices*. Edinburgh: Edinburgh University Press.
- Hanley, Richard (1997). *The Metaphysics of Star Trek*. New York: Basic Books.
- Hanson, Norwood Russell (1958). *Patterns of Discovery*. Cambridge: Cambridge University Press.
- Harris, Henry ed. (1995). *Identity: Essays Based on Herbert Spencer Lectures Given in the University of Oxford*. New York: Clarendon/Oxford Press.
- Hart, Ivor B. (1923). *Makers of Science: Mathematics, Physics, Astronomy*. London: Oxford University Press.
- Helm, Paul (1975). "Are 'Cambridge' Changes Non-Events?" *Analysis*, Vol. 35, No.4 (March 1975): 140-144.

Bibliography of Works Cited

- Hertzberg, Lars (1991). "Imagination and the Sense of Identity." In Cockburn (1991): 143-156.
- Hesse, Mary (1966). *Models and Analogies in Science*. Notre Dame: University of Notre Dame Press.
- Hill, Christopher S. (1997). "Imaginability, Conceivability, Possibility and the Mind- Body Problem." *Philosophical Studies*, Vol. 87, No. 1 (July 1997): 61-85.
- Hintikka, Jaakko (1999). "The Emperor's New Intuitions." *Journal of Philosophy*, Vol. 96, No. 3. (1999): 127-147.
- Hirsch, Eli (1982). *The Concept of Identity*. New York: Oxford University Press.
- Hirsch, Eli (1991). "Divided Minds." *Philosophical Review*, Vol. 1 (January 1991): 3-30.
- Hirschfeld, Lawrence and Susan Gelman, eds. (1994). *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York: Cambridge University Press.
- Hobson, R. Peter (1990). "On the Origins of Self and the Case of Autism." *Development and Psychopathology* 2, 163-181.
- Horowitz, Tamara and Gerald Massey, eds. (1991). *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield Publishers.
- Horowitz, Tamara (1998). "Philosophical Intuitions and Psychological Theory." *Ethics*, Vol. 108, No. 2 (January 1998): 367-385. A version of this paper also appears in DePaul and Ramsey (1998), pp. 143-160.
- Hughes, Christopher (1997). "Same-Kind Coincidence and the Ship of Theseus." *Mind*, Vol. 106, No. 421 (January 1997): 53-67.

Bibliography of Works Cited

- Hume, David (1739/1978). *A Treatise of Human Nature*. Ed. L.A. Selby-Bigge, Oxford: Clarendon Press.
- Humphreys, Paul (1993). "Seven Theses on Thought Experiments." In Earman et al (1993): 205-227.
- Hurley, Susan (1989). *Natural Reasons*. New York: Oxford.
- Irvine, Andrew D. (1991). "On the Nature of Thought Experiments in Scientific Reasoning" in Horowitz and Massey (1991): 149-165.
- Janis, Allen I. (1991). "Can Thought Experiments Fail?" In Horowitz and Massey (1991): 113-118.
- Johnston, Mark (1987a). "Human Beings." *Journal of Philosophy*, Vol. 84, No. 2 (February 1987): 59-83.
- Johnston, Mark (1987b). "Review of Shoemaker and Swinburne" (1984). *Philosophical Review*, Vol. 96, No. 1 (January 1987): 123-128.
- Johnston, Mark (1989a). "Fission and the Facts." In Tomberlin (1989): 369-397.
- Johnston, Mark (1989b). "Relativism and the Self." In Krausz (1989): 441-472.
- Johnston, Mark (1992). "Reasons and Reductionism." *Philosophical Review*, Vol. 101, No. 3 (July 1992): 589-618.
- Johnston, Mark (1997). "Human Concerns Without Superlative Selves." In Dancy (1997): 149-179.
- Kant, Immanuel (1785/1964). *Groundwork of the Metaphysic of Morals*. Trans. H.J. Paton. New York: Harper Torchbooks.

- Kant, Immanuel (1787/1929). *Critique of Pure Reason*. Trans. Norman Kemp Smith. London: Macmillan.
- Keil, Frank (1979). *Semantic and Conceptual Development: An Ontological Perspective*. Cambridge: Harvard University Press.
- Keil, Frank (1987). "Conceptual Development and Category Structure." In Neisser, ed. (1987): 175-200.
- Keil, Frank C. (1989). *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: Bradford/MIT.
- King, Peter (1991). "Mediaeval Thought Experiments: The Metamethodology of Mediaeval Science." In Horowitz and Massey (1991): 43-64.
- Kolak, Daniel (1993). "The Metaphysics and Metapsychology of Personal Identity: Why Thought Experiments Matter in Deciding Who We Are." *American Philosophical Quarterly*, Vol. 30, No. 1 (January 1993): 39-53.
- Komatsu, Lloyd K. (1992). "Recent Views of Conceptual Structure." *Psychological Bulletin*, Vol. 112, No. 3 (1992): 500-526.
- Korsgaard, Christine (1989). "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy and Public Affairs*, Vol. 18, No. 2 (Spring 1989): 101-132.
- Koymans, Alexandre (1939/1979). *Galileo Studies*. Trans. J. Mepham. Atlantic Highlands: Humanities Press.
- Koymans, Alexandre (1960). "Galileo's Treatise De Motu Gravium: The Use and Misuse of an Imaginary Experiment." Reprinted in Koymans (1968): 44-88.
- Koymans, Alexandre (1968). *Metaphysics and Measurement*. London: Chapman and Hall.

- Krausz, Michael, ed. (1989). *Relativism: Interpretation and Confrontation*. Notre Dame: University of Notre Dame Press.
- Kuhn, Thomas (1964). "A Function for Thought Experiments." Reprinted with new pagination in Kuhn (1977).
- Kuhn, Thomas (1970). *The Structure of Scientific Revolutions* (second edition). Chicago: University of Chicago Press.
- Kuhn, Thomas (1977). *The Essential Tension*. Chicago: University of Chicago Press.
- Kujundzic, Nebojsa (1992). "How Does the Laboratory of the Mind Work?" *Dialogue*, Vol. 32, No. 3 (Summer 1992): 573-578.
- Kujundzic, Nebojsa (1995). "Thought Experiments: Architecture and Economy of Thought." *The Journal of the British Society for Phenomenology*, Vol. 26, No. 1 (January 1995): 86-93.
- Leibniz, G.W. (1765/1981). *New Essays on Human Understanding*. Translated and edited by Peter Remnant and Jonathan Bennett. Cambridge: Cambridge University Press.
- Lewis, David (1976). "Survival and Identity." In Rorty (1976): 17-40. Reprinted with new pagination in Lewis (1983a): 55-72.
- Lewis, David (1983a). *Philosophical Papers: Volume I*. New York and Oxford: Oxford University Press.
- Lewis, David (1983b). "Postscripts to 'Survival and Identity.'" In Lewis (1983a): 73-77.
- Lipton, P. "Review of Horowitz and Massey, eds. *Thought Experiments in Science and Philosophy* (1991)." *Ratio* (new series) 6: 82-86.

Locke, John (1710/1975). *An Essay Concerning Human Understanding* (Edited with a forward by Peter H. Nidditch). Oxford: Clarendon Press.

Lowe, E.J. (1983). "On the Identity of Artifacts." *Journal of Philosophy*, Vol. 80 (April 1983): 220-231.

Lowe, E.J. (1998). "Entity, Identity and Unity." *Erkenntnis*, Vol. 48, Nos. 2-3 (1998): 191-208.

MacDonald, G.F., ed. (1979). *Perception and Identity: Essays Presented to A.J. Ayer and His Replies*. Ithaca, NY: Cornell University Press

Mach, Ernst (1926/1976). *Knowledge and Error* (fifth edition). Trans. Thomas McCormack and Paul Foulkes.

Mach, Ernst (1933/1960). *The Science of Mechanics* (ninth edition). Trans. Thomas McCormack.

Mackie, J. L. (1976). *Problems from Locke*. Oxford: Clarendon Press.

Maddy, Penelope (1979). "Is the Importance of Identity Derivative?" *Philosophical Studies*, Vol. 35 (Fall 1979): 151-170.

Madell, Geoffrey (1981). *The Identity of the Self*. Edinburgh: Edinburgh University Press.

Madell, Geoffrey (1991). "Personal Identity and the Idea of a Human Being." In Cockburn (1991): 127-142.

Markman, Ellen M. (1989). *Categorization and Naming in Children: Problems of Induction*. Cambridge, MA: MIT/Bradford.

Martin, Raymond (1993). "Real Values: Why the Wilkes-Donagan Prohibition is Mistaken." *Metaphilosophy*, Vol. 24, No. 4 (October 1993): 400-406.

Martin, Raymond (1995). "Fission Rejuvenation." *Philosophical Studies*, Vol. 80, No. 1 (October 1995): 17-40.

Martin, Raymond, John Barresi, Alessandro Giovannelli, eds. (1998). "Fission Examples in the Eighteenth and Early Nineteenth Century Personal Identity Debate." *History of Philosophy Quarterly*, Vol. 15, No. 3 (July 1998): 323-348.

Massey, Gerald (1991). "Backdoor Analyticity." In Horowitz and Massey (1991): 285-296.

Massey, Gerald (1995). "*Gedankenexperiment*: Where Science and Philosophy Meet." Unpublished manuscript.

McDowell, John (1994). *Mind and World*. Cambridge: Harvard University Press.

McDowell, John (1997). "Reductionism and the First Person." In Dancy (1997): 230-250.

McMullin, Ernan (1978). "The Conception of Science in Galileo's Work." In Butts and Pitt (1978): 209-257.

McMullin, Ernan (1985). "Galilean Idealization." *Studies in History and Philosophy of Science*, Vol. 16 (1985): 247-273.

Medin, Douglas, L. and Robert. L. Goldstone (1990). "Concepts." In Michael W. Eysenck, ed., *The Blackwell Dictionary of Cognitive Psychology*. Cambridge, MA: Blackwell.

Medin, Douglas L. and Edward E. Smith (1984). "Concepts and Concept Formation." *Annual Review of Psychology*, Vol. 35, (1984): 113-138.

- Merricks, Trenton (1997). "Fission and Personal Identity Over Time." *Philosophical Studies*, Vol. 88, No. 2 (November 1997): 163-186.
- Merricks, Trenton (1998a). "Against the Doctrine of Microphysical Supervenience." *Mind*, Vol. 107 (January 1998): 59-71.
- Merricks, Trenton (1998b). "There Are No Criteria of Identity Over Time." *Noûs*, Vol. 32, No. 1 (March 1998): 106-124.
- Merricks, Trenton (1998c). "On Whether Being Conscious Is Intrinsic." *Mind*, Vol. 107, No. 428 (October 1998): 845-846.
- Miller, Fred D. and Nicholas Smith, eds. (1989). *Thought Probes: Philosophy Through Science Fiction Literature*. Englewood Cliffs, NJ: Prentice Hall.
- Mills, Eugene (1993). "Dividing without Reducing: Bodily Fission and Personal Identity." *Mind*, Vol. 102, No. 405 (January 1993): 37-51.
- Molnar, G. (1969). "Kneale's Argument Revisited." *Philosophical Review*. Reprinted in T.L. Beauchamp ed. *Philosophical Problems of Causation*. Belmont CA: Dickenson, 1974.
- Munitz, Milton K., ed. (1971). *Identity and Individuation*. New York: New York University Press.
- Nagel, Thomas (1971). "Brain Bisection and the Unity of Consciousness." *Synthese*, Vol. 22. Reprinted with new pagination in Perry (1975a): 227-245.
- Neisser, Ulrich, ed. (1987). *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge: Cambridge University Press.

- Nersessian, Nancy (1984). *Faraday to Einstein: Constructing Meaning in Scientific Theories*. Dordrecht and Boston: Kluwer.
- Nersessian, Nancy (1992). "How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science." In Giere (1992): 3-44.
- Nersessian, Nancy (1993). "In the Theoretician's Laboratory: Thought Experiment as Mental Modeling." *Proceedings of the Philosophy of Science Association*, Vol. 2 (1993): 291-301.
- Noonan, Harold (1985). "Wiggins, Artifact Identity and 'Best Candidate' Theories." *Analysis*, Vol. 45 (January 1985): 4-8.
- Noonan, Harold (1986). "Reply to Garrett's 'Noonan, 'Best Candidate' Theories and the Ship of Theseus.'" *Analysis*, Vol. 46 (October 1986): 205-211.
- Noonan, Harold (1988). "Reply to Lowe on Ships and Structures." *Analysis*, Vol. 48 (October 1988): 221-223.
- Noonan, Harold (1989). *Personal Identity*. London and New York: Routledge.
- Norton, John (1991). "Thought Experiments in Einstein's Work." In Horowitz and Massey (1991): 129-148.
- Norton, John (1996). "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy*, Vol. 26, No. 3 (September 1996): 333-366.
- Nozick, Robert (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- Nozick, Robert (1981). *Philosophical Explanations*. Cambridge, MA: Belknap Press.

- Nozick, Robert (1993). *The Nature of Rationality*. Princeton, NJ: Princeton University Press.
- Oderberg, David S. *The Metaphysics of Identity over Time*. New York: St. Martin's Press.
- Oderberg, David S. (1989a). "Johnston on Human Beings." *Journal of Philosophy*, Vol. 86, No. 3 (March 1989): 137-141.
- Olson, David R. and Nancy Torrance, eds. (1993). *Modes of Thought: Explorations in Culture and Cognition*. Cambridge: Cambridge University Press.
- Olson, Eric (1997a). *The Human Animal: Personal Identity Without Psychology*. Oxford: Oxford University Press.
- Olson, Eric (1997b). "Relativism and Persistence." *Philosophical Studies*, Vol. 88, No. 2 (November 1997): 141-162.
- O'Neill, Onora (1986). "The Power of Example." *Philosophy*, Vol. 61 (1986): 5-29.
- Parfit, Derek (1971a). "On the Importance of Self-Identity." *Journal of Philosophy*, Vol. 68 (October 1971): 683-690.
- Parfit, Derek (1971b). "Personal Identity." *Philosophical Review*, Vol. 80, No. 1 (January, 1971). Reprinted with new pagination in Perry (1975a): 199-223.
- Parfit, Derek (1976). "Lewis, Perry and What Matters." In Rorty (1976): 91-107.
- Parfit, Derek (1984/1987). *Reasons and Persons*. Oxford: Clarendon Press.
- Parfit, Derek (1986). "Comments." *Ethics*, Vol. 96 (July 1986): 832-872.

- Parfit, Derek (1987a). "Divided Minds and the Nature of Persons." In Blakemore and Greenfield (1987).
- Parfit, Derek (1994). "Personal Identity." Unpublished manuscript.
- Parfit, Derek (1995). "The Unimportance of Identity." In Harris (1995): 13-47.
- Peirce, Charles Sanders (1867-1893/1992). *The Essential Peirce: Selected Philosophical Writings*. Vol. I. Ed. Nathan Houser and Christian Kloesel. Bloomington and Indianapolis: Indiana University Press.
- Penelhum, Terence (1957). "Is It Possible That One and the Same Individual Object Should Cease to Exist and Later On, Start to Exist Again?" *Analysis*, Vol. 17 (June 1957): 123-124.
- Penelhum, Terence (1970). *Survival and Disembodied Existence*. London: Routledge and Kegan Paul; New York: Humanities Press.
- Perry, John (1972). "Can the Self Divide?" *Journal of Philosophy*, Vol. 69, No. 16 (September 7, 1972): 463-488.
- Perry, John, ed. (1975a). *Personal Identity*. Berkeley and Los Angeles: University of California Press.
- Perry, John (1975b). "Personal Identity, Memory and the Problem of Circularity." In Perry (1975a): 135-155.
- Perry, John (1975c). "The Problem of Personal Identity." In Perry (1975a): 3-30.
- Perry, John (1976). "The Importance of Being Identical." In Rorty (1976): 67-90.

- Piaget, Jean (1946/1970). *The Child's Conception of Movement and Speed*. Trans. G.E.T. Holloway and M.J. Mackenzie. NY: Basic Books.
- Plantiga, Alvin (1974). *The Nature of Necessity*. Oxford: Oxford University Press.
- Plato (c. 380BC/1974). *Republic*. Trans. G.M.A. Grube revised C.D.C. Reeve. Indianapolis: Hackett.
- Poincaré, Henri (1905/1952). *Science and Hypothesis*. New York: Dover.
- Popper, Karl (1959/1992). *The Logic of Scientific Discovery*. Trans. Karl Popper. New York: Routledge.
- Prudovsky, Gad (1989). "The Confirmation of the Superposition Principle: The Role of a Constructive Thought Experiment in Galileo's Discorsi." *Studies in History and Philosophy of Science*, Vol. 20, No. 4 (December 1989): 453-468.
- Putnam, Hilary (1975a). "The Meaning of Meaning." In Putnam (1975b): 215-271.
- Putnam, Hilary (1975b). *Mind, Language and Reality: Philosophical Papers, Volume 2*. Cambridge: Cambridge University Press.
- Putnam, Hilary (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Putnam, Hilary (1990a). "Is Water Necessarily H₂O?" In Putnam (1990b): 54-79.
- Putnam, Hilary (1990b). *Realism with a Human Face*. Cambridge: Harvard University Press.
- Putnam, Hilary (1994a). "Rethinking Mathematical Necessity." In Putnam (1994b): 245-263.

- Putnam, Hilary (1994b). *Words and Life*. Cambridge: Harvard University Press.
- Quine, W.V. (1972). "Review of Identity and Individuation." *Journal of Philosophy*, Vol. 69, No. 16 (September 7, 1972): 488-497.
- Quinton, Anthony (1962/1975). "The Soul." *Journal of Philosophy*, Vol. 59, No. 15 (July 1962). Reprinted with new pagination in Perry (1975a): 53-72.
- Rea, Michael C. (1995). "The Problem of Material Constitution." *Philosophical Review*, Vol. 104, No. 4 (October 1995): 525-553.
- Rea, Michael C., ed. (1997) *Material Constitution: A Reader*. Savage, Maryland: Rowman & Littlefield.
- Rea, Michael C. (1998). "Sameness Without Identity: An Aristotelian Solution to the Problem of Material Constitution." *Ratio*, Vol. 11, No. 3 (December 1998): 316-328.
- Reid, Thomas (1785/1975). "Of Identity." Reprinted in Perry (1975a): 107-112.
- Reid, Thomas (1785/1975). "Of Mr. Locke's Account of Our Personal Identity." Reprinted in Perry (1975a): 113-118.
- Rescher, Nicholas (1991). "Thought Experiment in Pre-Socratic Philosophy." In Horowitz and Massey (1991): 31-41.
- Robinson, Denis (1982). "Reidentifying Matter." *Philosophical Review*, Vol. 91 (1982): 317-341.
- Robinson, Denis (1985). "Can Amoeba Divide without Multiplying?" *Australasian Journal of Philosophy*, Vol. 63: 299-319.

- Rorty, Amélie Oksenberg, ed. (1976). *The Identities of Persons*. Berkeley: University of California Press.
- Rosenberg, Jay F. and Charles Travis, eds. (1971). *Readings in the Philosophy of Language*. Englewood Cliffs, NJ: Prentice Hall.
- Rosenberg, Jay F. (1993). "Comments on Peter van Inwagen's *Material Beings*." *Philosophy and Phenomenological Research*, Vol. 53, No. 3 (September 1993): 701-708.
- Rosch, Eleanor and C. G. Mervis (1975). "Family Resemblance: Studies in the Internal Structure of Categories." *Cognitive Psychology*, Vol. 7: 573-605.
- Rosch, Eleanor and Lloyd B. B. (1978). *Cognition and Categorization*. Hillsdale, NJ: Erlbaum Associates.
- Rovane, Carol (1990). "Branching Self-Consciousness." *Philosophical Review*, Vol. 99, No. 3 (July 1990): 355-395.
- Rovane, Carol (1993). "Self-Reference: The Radicalization of Locke." *Journal of Philosophy*, Vol. 90, No. 2 (February 1993): 73-97.
- Rovane, Carol (1998). *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton, NJ: Princeton University Press.
- Rubiner , Michael (1996). "Endpaper: Greetings." *New York Times Magazine* (January 14, 1996): 60.
- Salmon, Nathan (1982). *Reference and Essence*. Oxford: Basil Blackwell.
- Scaltsas, Theodore (1980). "The Ship of Theseus." *Analysis*, Vol. 40, No. 3 (June 1980): 152-157.

- Schechtman, Marya (1990). "Personhood and Personal Identity." *Journal of Philosophy*, Vol. 77, No. 2 (February 1990): 71-92.
- Schechtman, Marya (1996). *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
- Sedley, David (1982). "The Stoic Criterion of Identity." *Phronesis*, Vol. 22 (1982): 255-275.
- Segre, M. (1989). "Galileo, Viviani and the Tower of Pisa." *Studies in History and Philosophy of Science*, 20: 453-468
- Shanks, Niall, ed. (1998). *Idealization IX: Idealization in Contemporary Physics*. Amsterdam: Rodopi.
- Shepard, Roger (1994). "Mind and World: From Newton, Einstein and Darwin to Principles of Mind." Unpublished *William James Lectures*, Harvard University, Fall 1994.
- Shoemaker, Sydney (1959/1975). "Personal Identity and Memory." *Journal of Philosophy*, Vol. 56, No. 22 (October 22, 1959). Reprinted with new pagination in Perry (1975a): 119-134.
- Shoemaker, Sydney (1963). *Self-Knowledge and Self-Identity*. Ithaca, NY: Cornell University Press.
- Shoemaker, Sydney (1976). "Embodiment and Behavior." In Rorty (1976): 109-137.
- Shoemaker, Sydney (1984). *Identity, Cause and Mind*. Cambridge: Cambridge University Press.
- Shoemaker, Sydney (1985). "Critical Notice of Reasons and Persons." *Mind*, Vol. 94 (1985): 443-453

- Shoemaker, Sydney (1994). "The First-Person Perspective." *Proceedings of the APA*, Vol. 68, No. 2 (November 1994): 7-22.
- Shoemaker, Sydney (1997a). "Parfit on Identity." In Dancy (1997): 135-148.
- Shoemaker, Sydney (1997b). "Self and Substance." In Tomberlin (1997).
- Shoemaker, Sydney and Richard Swinburne (1984). *Personal Identity: Great Debates in Philosophy*. Oxford: Basil Blackwell.
- Smart, Brian (1972). "How to Reidentify the Ship of Theseus." *Analysis* Vol. 32, No. 5 (April 1972): 145-148.
- Smart, Brian (1973). "The Ship of Theseus, the Parthenon, and Disassembled Objects." *Analysis*, Vol. 34, No. 1 (October 1973): 24-27.
- Smith, Edward E. and Douglas L. Medin (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.
- Snowdon, Paul F. (1991). "Personal Identity and Brain Transplants." In Cockburn (1991): 109-126.
- Sorensen, Roy (1992a). *Thought Experiments*. New York and Oxford: Oxford University Press.
- Sorensen, Roy (1992b). "Thought Experiments and the Epistemology of Laws." *Canadian Journal of Philosophy*, Vol. 22, No. 1 (March 1992), 15-44.
- Sosa, Ernest (1990). "Surviving Matters." *Noûs*, Vol. 24, No. 2 (April 1990): 305-330.
- Stevin, Simon (1955). *The Principal Works of Simon Stevin. Vol. 1: General Introduction and Mechanics*. Ed. E.J. Dijksterhuis. Amsterdam: C. W. Swets and Zeitlinger.

- Stone, Tony and Martin Davies, eds. (1995). *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.
- Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.
- Sunstein, Cass (1993). "On Analogical Reasoning." *Harvard Law Review*, Vol. 106: 741-791.
- Sunstein, Cass (1996). "Political Conflict and Legal Agreement." In Grethe (1996).
- Taylor, Charles (1977). "Self-Interpreting Animals." Reprinted in Taylor (1985), 45-76.
- Taylor, Charles (1981). "The Concept of a Person." Reprinted in Taylor (1985), 97-114.
- Taylor, Charles (1985). *Human Agency and Language: Philosophical Papers, Volume 1*. Cambridge: Cambridge University Press.
- Thomason, Sarah (1991). "Thought Experiments in Linguistics." In Horowitz and Massey (1991): 247-257.
- Thomson, Judith Jarvis (1971). "A Defense of Abortion." *Philosophy and Public Affairs*, Vol. 1, No. 1 (Fall 1971): 47-66.
- Thomson, Judith Jarvis (1986). *Rights, Restitution and Risk*. Cambridge: Harvard University Press.
- Thomson, Judith Jarvis, ed. (1987a). *On Being and Saying: Essays in Honor of Richard Cartwright*. Cambridge, MA: MIT Press.
- Thomson, Judith Jarvis, ed. (1987b). "Ruminations on an Account of Personal Identity." In Thomson (1987a): 215-240.

- Thomson, Judith Jarvis (1990). *The Realm of Rights*. Cambridge: Harvard University Press.
- Thomson, Judith Jarvis (1997). "People and Their Bodies." In Dancy (1997): 202-229.
- Tidman, Paul (1994). "Conceivability as a Test for Possibility." *American Philosophical Quarterly*, Vol. 31, No. 4 (October 1994): 297-309.
- Tomberlin, James E., ed. (1989) *Philosophical Perspectives*, 3. Atascadero, CA: Ridgeview Publishing Co.
- Tomberlin, James E., ed. (1997). *Mind, Causation and World*. Boston, MA: Blackwell.
- Unger, Peter (1979a). "I Do Not Exist." In MacDonald (1979).
- Unger, Peter (1979b). "Why There Are No People." *Midwest Studies in Philosophy*, Vol. 4 (1979): 177-222.
- Unger, Peter (1990). *Identity, Consciousness and Value*. New York and Oxford: Oxford University Press
- Van Inwagen, Peter (1980a). "Philosophers and the Words 'Human Body'." In van Inwagen (1980b): 283-300.
- Van Inwagen, Peter, ed. (1980b). *Time and Cause: Essays Presented to Richard Taylor*. Dordrecht: Reidel.
- Van Inwagen, Peter (1990). *Material Beings*. Ithaca, NY: Cornell University Press.
- Van Inwagen, Peter (1993a). "Critical Study of Peter Unger's Identity Consciousness and Value." *Noûs*, Vol. 27, No. 2: 373-379
- Velleman, J. David (1996). "Self to Self." *Philosophical Review*, Vol. 105, No. 1 (1996): 39-76.

- Walton, Kendall (1990). *Mimesis as Make-Believe*. Cambridge: Harvard University Press.
- Wellman, H. M. (1990). *The Child's Theory of Mind*. Cambridge: MIT Press.
- White, Stephen (1989). "Metapsychological Relativism and the Self." *Journal of Philosophy*, Vol. 86 (1989): 298-322.
- White, Stephen (1995). "The Desire to Survive." *Philosophy and Phenomenological Research*, Vol. 52, No. 1 (March 1992): 153-158.
- Whiting, Jennifer (1986). "Friends and Future Selves." *Philosophical Review*, Vol. 95, No. 4 (October 1986): 547-580.
- Wiggins, David (1967). *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.
- Wiggins, David (1968). "On Being in the Same Place at the Same Time." *Philosophical Review*, Vol. 77 (1968): 90-95.
- Wiggins, David (1979). "The Concern to Survive." *Midwest Studies in Philosophy*, Vol. 4 (1979): 417-422.
- Wiggins, David (1980). *Sameness and Substance*. Cambridge, MA: Harvard University Press.
- Wilkes, Kathleen V. (1988). *Real People: Personal Identity without Thought Experiments*. Oxford: Clarendon Press.
- Williams, Bernard (1956/1973). "Personal Identity and Individuation." Reprinted (with new pagination) in Williams (1973): 1-18.
- Williams, Bernard (1960/1973). "Bodily Continuity and Personal Identity." Reprinted (with new pagination) in Williams (1973): 19-25.

- Williams, Bernard (1966/1973). "Imagination and the Self." Reprinted (with new pagination) in Williams (1973): 26-45.
- Williams, Bernard (1970a/1973). "Are Persons Bodies?" Reprinted (with new pagination) in Williams (1973): 64-81.
- Williams, Bernard (1970b/1973). "The Self and the Future." Reprinted (with new pagination) in Williams (1973): 46-63. Also in Perry (1975a): 179-198.
- Williams, Bernard (1973). *Problems of the Self*. Cambridge: Cambridge University Press.
- Wiser, Marianne and Susan Carey (1983). "When Heat and Temperature Were One." In Gentner and Stevens (1983): 267-297.
- Witt-Hansen, Johannes (1976). "H.C. Orsted, Immanuel Kant, and the Thought Experiment." *Danish Yearbook of Philosophy*, Vol. 13 (1976): 48-65.
- Wittgenstein, Ludwig (1953). *Philosophical Investigations*. Trans. G.E.M. Anscombe. Oxford: Basil Blackwell.
- Wolf, Susan (1986). "Self-Interest and Interest in Selves." *Ethics*, Vol. 96 (July 1986): 704-720.
- Yablo, Stephen (1993). "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research*, Vol. 53, No.1 (March 1993): 1-42.

6.2 BIBLIOGRAPHY ON GALILEO, EXPERIMENT AND THOUGHT EXPERIMENT

(Prepared with the assistance of David Bzdak)

Note: This is not meant to be a comprehensive Galileo bibliography. Rather, it is a listing of books and articles which address topics related to Galileo and (thought) experimentation – especially, those related to Galileo's famous (thought) experiment concerning falling bodies.

- Akhutin, A.V. (1982). "Galileo's Experiment." *Scientia*, Vol. 117 (1982): 287-312.
- Brown, James Robert (1991a). *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. New York and London: Routledge.
- Brown, James Robert (1991b). "Thought Experiments: A Platonic Account." In Horowitz and Massey (1991): 119-128.
- Brown, James Robert (1993a). "Author's Response (to Norton [1993])." *Metascience* (new series), Vol. 3 (1993): 38-40.
- Brown, James Robert (1993b). "Why Empiricism Won't Work." *Proceedings of the Philosophy of Science Association* 2, 271-279.
- Butts, Robert E. and Joseph C. Pitt, eds. (1978). *New Perspectives on Galileo*. Dordrecht/ Boston: Reidel.
- Cargile, James (1987). "Definitions and Counterexamples." *Philosophy*, Vol. 62: 179-193.
- Casper, B.M. (1977). "Galileo and the Fall of Aristotle: A Case of Historical Injustice?" *American Journal of Physics*, Vol. 45 (1977): 325-330

Claggett, Marshall (1959). *The Science of Mechanics in the Middle Ages*. Madison: University of Wisconsin Press.

Clavelin, Maurice (1974). *The Natural Philosophy of Galileo: Essays on the Origins and Formation of Classical Mechanics*. Trans. A. J. Pomerans. Cambridge: MIT Press.

Clement, John (1983). "A Conceptual Model Discussed by Galileo and Used Intuitively by Physics Students." In Gentner and Stevens (1983): 325-340.

Cooper, Lane (1935). *Aristotle, Galileo, and the Tower of Pisa*. Ithaca, NY: Cornell University Press.

Damerow, Peter et al. (1992). *Exploring the Limits of Pre-Classical Mechanics: A Study of Conceptual Development in Early Modern Science: Free Fall and Compounded Motion in the Work of Descartes, Galileo and Beeckman*. New York: Springer Verlag.

Dijksterhuis, E. J. (1961, repr. 1986). *The Mechanization of the World Picture: Pythagoras to Newton*. Trans. C. Dikshoorn. Princeton, NJ: Princeton University Press.

Drake, Stillman (1977). "Galileo and the Career of Philosophy." *Journal of the History of Philosophy*, Vol. 38 (January-March 1977): 19-32.

Drake, Stillman (1978). *Galileo at Work: His Scientific Biography*. Chicago, IL: University of Chicago Press.

Drake, Stillman (1989). *History of Free Fall: Aristotle to Galileo*. Toronto: Wall & Thompson.

Drake, Stillman (1990). *Galileo: Pioneer Scientist*. Toronto: University of Toronto Press.

Drake, Stillman and I. E. Drabkin (1969). *Mechanics in Sixteenth Century Italy: Selections from Tartaglia, Benedetti, Guido Ubaldo, Galileo*. Madison: University of Wisconsin Press.

Finocchiaro, Maurice A. (1994). "Methodological Judgment and Critical Reasoning in Galileo's Dialogue." *Proceedings of the Biennial Meetings of the Philosophy of Science Association*, Vol. 2 (1994): 248-257.

Galilei, Galileo (1590/1960). *On Motion*. Madison, WI: University of Wisconsin Press, trans. I. E. Drabkin.

Galilei, Galileo (1632/1967). *Dialogue Concerning the Two Chief World Systems*. Trans. Stillman Drake. Berkeley: University of California Press.

Galilei, Galileo (1638/1914, repr. 1954). *Dialogues Concerning Two New Sciences*. Trans. Henry Crew and Alfonso de Salvio. New York: Dover.

Galilei, Galileo (1638/1974, rev. 1989). *Two New Sciences, Including Centers of Gravity and Force of Gravity and Force of Percussion*. Trans. Stillman Drake. Toronto: Wall & Thompson.

Gendler, Tamar Szabó (1998). "Galileo and the Indispensability of Scientific Thought Experiment." *British Journal for the Philosophy of Science*, Vol. 49, No. 3 (September 1998): 397-424.

Giere, R. ed. (1992). *Cognitive Models of Science*. Minneapolis: University of Minnesota Press.

Gooding, David C., Pinch, T. and Schaffer, S. (1989). *The Uses of Experiment: Studies in the Natural Sciences*. Cambridge: Cambridge University Press.

Hart, Ivor B. (1923). *Makers of Science: Mathematics, Physics, Astronomy*. London: Oxford University Press.

Koertge, Noretta (1977). "Galileo and the Problem of Accidents." *Journal of the History of Ideas*, Vol. 38 (July-September 1997): 389-408.

Koyré, Alexandre (1939/1979). *Galileo Studies*. Trans. J. Mepham. Atlantic Highlands: Humanities Press.

Koyré, Alexandre (1960). "Galileo's Treatise De Motu Gravium: The Use and Misuse of an Imaginary Experiment." Reprinted in Koyré (1968): 44-88.

Koyré, Alexandre (1968). *Metaphysics and Measurement*. London: Chapman and Hall.

Kuhn, Thomas (1964). "A Function for Thought Experiments." Reprinted with new pagination in Kuhn (1977).

Kuhn, Thomas (1970). *The Structure of Scientific Revolutions* (second edition). Chicago: University of Chicago Press.

Kuhn, Thomas (1977). *The Essential Tension*. Chicago: University of Chicago Press.

Kujundzic, Nebojsa (1992). "How Does the Laboratory of the Mind Work?" *Dialogue*, Vol. 32, No. 3 (Summer 1992): 573-578.

Machamer, Peter, ed. (1998). *The Cambridge Companion to Galileo*. New York: Cambridge University Press.

MacLachlan, J. (1973). "A Test of an 'Imaginary' Experiment of Galileo's." *Isis*, Vol. 64 (1973): 374-379.

McMullin, Ernan (1978). "The Conception of Science in Galileo's Work." In Butts and Pitt (1978): 209-257.

McMullin, Ernan (1985). "Galilean Idealization." *Studies in History and Philosophy of Science*, Vol. 16 (1985): 247-273.

Naylor, R. H. (1989). "Galileo's Experimental Discourse" in Gooding, Pinch and Schaffer (1989): 117-134.

Norton, John (1996). "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy*, Vol. 26, No. 3 (September 1996): 333-366.

Prudovsky, Gad (1989). "The Confirmation of the Superposition Principle: The Role of a Constructive Thought Experiment in Galileo's Discorsi." *Studies in History and Philosophy of Science*, Vol. 20, No. 4 (December 1989): 453-468.

Segre, M. (1980). "The Role of Experiment in Galileo's Physics." *Archive for History of Exact Sciences*, Vol. 23 (1980): 227-252.

Segre, M. (1989). "Galileo, Viviani and the Tower of Pisa." *Studies in History and Philosophy of Science*, 20: 453-468.

Shapin, S. and Schaffer, S. (1985). *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton, NJ: Princeton University Press.

Shea, W. R. (1972). *Galileo's Intellectual Revolution*. London: Macmillan.

Shepard, Roger (1994). "Mind and World: From Newton, Einstein and Darwin to Principles of Mind." Unpublished William James Lectures, Harvard University, Fall 1994.

6.3 BIBLIOGRAPHY ON PERSONAL IDENTITY AND IDENTITY

(Prepared by David Bzdak)

Alexander, Ronald G. (1997). *Self, Supervenience and Personal Identity*. Hampshire, UK: Ashgate Publishing Com.

Alston, William and Jonathan Bennett (1988). "Locke on People and Substances." *Philosophical Review*, Vol. 97 (1988): 25-46.

Apter, Andrew (1991). "The Problem of Who: Multiple Personality, Personal Identity, and the Double Brain." *Philosophical Psychology*, Vol. 4 (1991): 219-248.

Aune, Bruce (1994). "Speaking of Selves." *Philosophical Quarterly*, Vol. 44, No. 176 (July 1994): 279-293.

Baillie, James (1990). "Identity, Survival, and Sortal Concepts." *Philosophical Quarterly*, Vol. 40 (April 1990): 183-194.

Baillie, James (1991). "Split Brains and Single Minds." *Journal of Philosophical Research*, Vol. 16 (1991): 11-18.

Baillie, James (1993a). *Problems in Personal Identity*. New York: Paragon House.

Baillie, James (1993b). "Recent Work on Personal Identity." *Philosophical Books*, Vol. 34, No. 4 (October, 1993): 193-206.

Baillie, James (1993c). "What Matters in Survival?" *Southern Journal of Philosophy*, Vol. 31, No. 3 (Fall 1993): 255-261.

Baillie, James (1996). "Identity Relation R, and What Matters." *Pacific Philosophical Quarterly*, Vol. 77, No. 4 (December 1996): 263-267.

Baillie, James (1997). "Personal Identity and Mental Content." *Philosophical Psychology*, Vol. 10, No. 3 (September 1997): 323-333.

Beck, Simon (1989). "Parfit and the Russians (Personal Identity and Moral Concepts)." *Analysis*, Vol. 49 (1989): 205-209.

Beck, Simon (1992). "Should We Tolerate People Who Split?" *The Southern Journal of Philosophy*, Vol. 30, No. 1 (Spring 1992): 1-17.

Beck, Simon (1998). "Back to the Self and the Future." *South African Journal of Philosophy*, Vol. 17, No. 3 (August 1998): 211-225.

Belzer, Marvin (1996). "Notes on Relation R." *Analysis*, Vol. 56, No. 1 (January 1996): 56-62.

Bermúdez, José, Anthony Marcel and Naomi Eilan, eds. (1998). *The Body and the Self*. Cambridge, MA: MIT Press.

Bermúdez, José (1998). *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.

Blackburn, Simon (1997). "Has Kant Refuted Parfit?" In Dancy (1997): 180-201.

Blakemore, Colin and Susan Greenfield, eds. (1987). *Mindwaves*. Oxford: Blackwell.

Bodansky, E. (1987). "Parfit on Selves and their Interests." *Analysis*, Vol. 47 (1987): 47-50.

Bourgeois, Warren (1995). *Persons: What Philosophers Say About You*. Waterloo, ON: Wilfrid Laurier University Press.

Braude, Stephen E. (1991). *First-Person Plural: Multiple Personality and the Philosophy of Mind*. New York and London: Routledge.

Brennan, Andrew (1982). "Personal Identity and Personal Survival." *Analysis*, Vol. 42 (January 1982): 44-50.

Brennan, Andrew (1984). "Survival." *Synthese*, Vol. 59 (June 1984): 39-62.

Brennan, Andrew (1987a). "Discontinuity and Identity." *Noûs*, Vol. 21 (June 1987): 241-260.

Brennan, Andrew (1987b). "Survival and Importance." *Analysis*, Vol. 47 (October 1987): 225-230.

Brennan, Andrew (1988). *Conditions of Identity: A Study of Identity and Survival*. Oxford: Oxford University Press.

Brody, Baruch (1972). "Locke on the Identity of Persons." *American Philosophical Quarterly*, Vol. 9 (October 1972): 327-334.

Brody, Baruch (1980). *Identity and Essence*. Princeton: Princeton University Press.

Broome, J. (1984). "Indefiniteness of Identity." *Analysis*, Vol. 44 (1984): 6-12.

Broyles, James (1986). "Wittgenstein On Personal Identity." *Philosophical Investigations*, Vol. 9 (January 1986): 56-65.

Brueckner, Anthony (1993). "Parfit on What Matters in Survival." *Philosophical Studies*, Vol. 70, No. 1 (April 1993): 1-22.

Burke, Michael B (1980). "Cohabitation, Stuff and Intermittent Existence." *Mind*, Vol. 89 (July 1980): 391-405.

Burke, Michael B (1983). "Identity and Origin." *Dialogos*, Vol. 18 (April 1993): 59-66.

Burke, Michael B (1992). "Copper Statues and Pieces of Copper: A Challenge to the Standard Account." *Analysis*, Vol. 52, No. 1 (January 1992): 12-17.

Burke, Michael B. (1994a). "Dion and Theon: An Essentialist Solution to an Ancient Puzzle." *Journal of Philosophy*, Vol. 91, No. 3 (March 1994): 129-139.

Burke, Michael B. (1994b). "Preserving the Principle of One Object to a Place: A Novel Account of the Relations Among Objects, Sorts, Sortals, and Persistence Conditions." *Philosophy and Phenomenological Research*, Vol. 54, No. 3 (September 1994): 591-624.

Burke, Michael B. (1997a). "Persons and Bodies: How to Avoid the New Dualism." *American Philosophical Quarterly*, Vol. 34, No. 4 (October 1997): 457-467.

Burke, Michael B. (1997b). "Reply to Lowe and Denkel." *Analysis*, Vol. 57, No. 1 (January 1997): 11-18.

Bushnell, Dana E. (1993). "Identity, Psychological Continuity, and Rationality." *Journal of Philosophical Research*, Vol. 18 (1993): 15-24.

Butler, Joseph (1736/1975). "Of Personal Identity." *First Appendix to The Analogy of Religion*. Reprinted with new pagination in Perry 1975a.

Care, Norman S. and Robert H. Grimm, eds. (1969). *Perception and Personal Identity: Proceedings*. Cleveland, OH: Case Western University Press.

Carrithers, Michael, Steven Collins and Steven Lukes, eds. (1985). *The Category of the Person: Anthropology, Philosophy, History*. Cambridge: Cambridge University Press.

Cartwright, Helen M. (1993). "On Two Arguments for the Indeterminacy of Personal Identity." *Synthese*, Vol. 95, No. 2 (May 1993): 241-273.

Cassam, Quassim (1993). "Parfit on Persons." *Proceedings of the Aristotelian Society*, Vol. 93 (1993): 17-37.

Cassam, Quassim (1997). *Self and World*. Oxford: Clarendon/Oxford Press.

Chalmers, David (undated). University of Arizona. On-Line, Personal Identity Bibliography. <http://www.u.arizona.edu/~chalmers/biblio/3.html#3.7>

Chandra, Suresh (1977). *Identity and Thought Experiment*. Simla: Indian Institute for Advanced Study.

Chappel, T. (1995). "Personal Identity, R-Relatedness, and the Empty Question Argument." *Philosophical Quarterly*, Vol. 45, No. 178 (January 1995): 88 -92.

Chisholm, Roderick M (1971). "Problems of Identity." In Munitz (1971): 3-30.

Chisholm, Roderick M. (1976). *Person and Object*. London: Allen and Unwin.

Coburn, Robert C. (1960). "Bodily Continuity and Personal Identity." *Analysis*, Vol. 20 (April 1960): 117-120.

Cockburn, David, ed. (1991). *Human Beings*. Cambridge: Cambridge University Press.

Collins, Arthur W. (1997). "Personal Identity and the Coherence of Q-Memory." *Philosophical Quarterly*, Vol. 47, No. 186 (January 1997): 73-80.

Cowley, Fraser (1971). "The Identity of a Person and His Body." *Journal of Philosophy*, Vol. 68 (October 1971): 678-683.

Crabbe, J., ed. (1999). *From Soul to Self*. London: Routledge.

Curzer, Howard (1991). "An Ambiguity in Parfit's Theory of Personal Identity." *Ratio*, Vol. 4 (June 1991): 16-24.

Cuypers, Stefaan E. (1998). "Philosophical Atomism and the Metaphysics of Personal Identity." *International Philosophical Quarterly*, Vol. 38, No. 4 (December 1998): 349-368.

Dainton, Barry F. (1996). "Survival and Experience." *Proceedings of the Aristotelian Society*, Vol. 96 (1996): 17-36.

Dancy, Jonathan (1997). *Reading Parfit*. Oxford: Basil Blackwell.

Dauer, Francis W. (1972). "How Not to Reidentify the Parthenon." *Analysis*, Vol. 33, No. 2 (December 1972): 63-64.

Davis, Lawrence H. (1973). "Smart on Conditions of Identity." *Analysis*, 33:3 (January 1973): 109-110.

Davis, Lawrence H. (1997). "Cerebral Hemispheres." *Philosophical Studies*, Vol. 87, No. 2 (August 1997): 207-222.

Davis, Lawrence H. (1998). "Functionalism and Personal Identity." *Philosophy and Phenomenological Research*, Vol. 58, No. 4 (December 1998): 781-804.

Dennett, Daniel (1976). "Conditions of Personhood." In Rorty (1976): 175-196.

Dennett, Daniel (1989). "The Origins of Selves." *Cogito*, Vol. 3 (1989): 163-173.

- Doepke, Frederick (1990). "The Practical Importance of Personal Identity." *Logos (USA)*, (1990): 83-91.
- Doepke, Frederick (1996). *Kinds of Things: A Theory of Transcendental Argument*. Chicago, IL: Open Court Publishing Co.
- Donagan, Alan (1990). "Real Human Persons." *Logos (USA)*, Vol. 11 (1990): 1-16.
- Ehring, Douglas (1978). "Strawson and the Body Theory." *Dialogue (PST)*, Vol. 20 (April 1978): 33-41.
- Ehring, Douglas (1985). "Personal Identity and the Causal Theory of Memory." *Modern Schoolman*, Vol. 63 (November 1985): 65-69.
- Ehring, Douglas (1987a). "Personal Identity and Time Travel." *Philosophical Studies*, Vol. 52 (November 1987): 427-433.
- Ehring, Douglas (1987b). "Survival and Trivial Facts." *Analysis*, Vol. 47 (January 1987): 50-54.
- Ehring, Douglas (1995). "Personal Identity and the 'R'-Relation: Reconciliation Through Cohabitation?" *Australasian Journal of Philosophy*, Vol. 73, No. 3 (September 1995): 337-346.
- Eilan, Naomi (1998). "Consciousness and the Self." In Bermúdez, Marcel and Eilan (1998): 337-357.
- Elliot, Robert (1978). "Personal Identity, Reduplication and Spatio-Temporal Continuity." *Philosophical Papers*, Vol. 7 (October 1978): 73-75.
- Elliot, Robert (1981). "How to Travel Faster Than Light." *Analysis*, Vol. 41 (January 1981): 4-6.

- Elliot, Robert (1991). "Personal Identity and the Causal Continuity Requirement." *Philosophical Quarterly*, Vol. 41 (January 1991): 55-75.
- Epstein, Farryl (1973). "Metaphysics of Mind and Body Identity." *American Philosophical Quarterly*, Vol. 10 (1973): 111-121.
- Fields, Lloyd (1987). "Parfit on Personal Identity and Desert." *Philosophical Quarterly*, Vol. 37 (October 1987): 432-441.
- Frankfurt, Harry (1976). "Identification and Externality." In Rorty (1976): 239-252.
- French, P. T. Uehling and H. Wettstein, eds. (1984) *Midwest Studies in Philosophy*, IX. Minneapolis: University of Minnesota Press.
- Gale, Richard (1969) "A Note on Personal Identity and Bodily Continuity." *Analysis*, Vol. 29 (June 1969): 193-195.
- Gallagher, Shaun and Jonathan Shear, eds. (1999). *Models of the Self*. Exeter, UK: Imprint Academic.
- Gallagher, Shaun (undated). Canisius University, Buffalo. On-Line, Personal Identity Bibliography.
<http://gort.canisius.edu/gallaghr/>
- Garrett, Brian (1985). "Noonan, 'Best Candidate' Theories and the Ship of Theseus." *Analysis*, Vol. 45 (October 1985): 212-215.
- Garrett, Brian (1988). "Identity and Extrinsicsness." *Mind*, Vol. 97 (January 1988): 105-109.
- Garrett, Brian (1990a). "Personal Identity and Extrinsicsness." *Philosophical Studies*, Vol. 59, No. 2 (June 1990): 177-194.

- Garrett, Brian (1990b). "Persons and Human Beings." *Logos (USA)*, (1990): 47-56.
- Garrett, Brian (1991). "Personal Identity and Reductionism." *Philosophy and Phenomenological Research*, Vol. LI, No. 2 (June 1991): 361-373.
- Garrett, Brian (1998). *Personal Identity and Self-Consciousness*. London and New York: Routledge.
- Gendler, Tamar Szabó (1998). "Exceptional Persons: On the Limits of Imaginary Cases." *Journal of Consciousness Studies*, Vol., 5, Nos. 5-6 (1998): 592-610. Reprinted in Gallagher and Shear (1999): 447-466.
- Gendler, Tamar Szabó (1999). "Review of Eric Olson's The Human Animal." *Philosophical Review*, Vol. 108, No. 1 (January 1999): 112-116.
- Giles, James (1997). *No Self to be Found: The Search for Personal Identity*. New York: University Press of America.
- Gill, Christopher, ed. (1990). *The Person and the Human Mind*. Oxford: Clarendon Press.
- Gill, J.H. (1980). "Of Split Brains and Tacit Knowing." *International Philosophical Quarterly*, Vol. 20 (1980): 49-58.
- Gillett, Grant (1986). "Brain Bisection and Personal Identity." *Mind*, Vol. 95 (April 1986): 224-229.
- Glover, Jonathan (1988). *I: The Philosophy and Psychology of Personal Identity*. New York: Penguin.
- Goodenough, J.M. (1996) "Parfit and the Sorites Paradox." *Philosophical Studies*, Vol. 83, No. 2 (August 1996): 113-120.

- Greenwood, J.D. (1993). "Split Brains and Singular Personhood." *Southern Journal of Philosophy*, Vol. 31 (1993): 285-306.
- Grice, H. P. (1941/1975). "Personal Identity." *Mind*, Vol. 50 (October 1941). Reprinted with new pagination in Perry (1975a): 73-98.
- Hacking, Ian (1995). *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton, NJ: Princeton University Press.
- Haksar, Vinit (1991). *Indivisible Selves and Moral Practices*. Edinburgh: Edinburgh University Press.
- Hamilton, Andy (1995). "A New Look at Personal Identity." *Philosophical Quarterly*, Vol. 45, No. 180 (July 1995): 332-349.
- Hanley, Richard (1997). *The Metaphysics of Star Trek*. New York: BasicBooks.
- Harris, Henry (1995a). "An Experimentalist Looks at Identity." In Harris (1995b): 47-65.
- Harris, Henry ed. (1995b). *Identity: Essays Based on Herbert Spencer Lectures Given in the University of Oxford*. New York: Clarendon/Oxford Press.
- Haugen, David (1995). "Personal Identity and Concern for the Future." *Philosophia (Israel)*, Vol. 24, Nos. 3-4 (December 1995): 481-492.
- Hertzberg, Lars (1991). "Imagination and the Sense of Identity." In Cockburn (1991): 143-156.
- Hill, Christopher S. (1997). "Imaginability, Conceivability, Possibility and the Mind- Body Problem." *Philosophical Studies*, Vol. 87, No. 1 (July 1997): 61-85.

- Hirsch, Eli (1982). *The Concept of Identity*. New York: Oxford University Press.
- Hirsch, Eli (1991). "Divided Minds." *Philosophical Review*, Vol. 1 (January 1991): 3-30.
- Hirsch, Eli (1993). *Dividing Reality*. Oxford: Oxford University Press.
- Hobson, R. Peter (1990). "On the Origins of Self and the Case of Autism." *Development and Psychopathology* 2, 163-181.
- Hope, Tony (1994). "Personal Identity and Psychiatric Illness." *Philosophy*, Vol. 37 (1994 Supplement): 131-143.
- Huang, Yih-mei (1995). "Personal Identity and Thought Experiments." *Philosophical Review (Taiwan)*, Vol. 18 (January 1995): 93-110.
- Hughes, Christopher (1997). "Same-Kind Coincidence and the Ship of Theseus." *Mind*, Vol. 106, No. 421 (January 1997): 53-67.
- Humphrey, N. and Daniel Dennett (1989). "Speaking for Ourselves." *Raritan*, Vol. 9 (1989): 68-98.
- Johnston, Mark (1987a). "Human Beings." *Journal of Philosophy*, Vol. 84, No. 2 (February 1987): 59-83.
- Johnston, Mark (1987b). Review of Shoemaker and Swinburne (1984). *Philosophical Review*, Vol. 96, No. 1 (January 1987): 123-128.
- Johnston, Mark (1989a). "Fission and the Facts." In Tomberlin (1989): 369-397.
- Johnston, Mark (1989b). "Relativism and the Self." In Krausz (1989): 441-472.

- Johnston, Mark (1992a). "Constitution is Not Identity." *Mind*, Vol. 101, No. 401 (January 1992): 89-105.
- Johnston, Mark (1992b). "Reasons and Reductionism." *Philosophical Review*, Vol. 101, No. 3 (July 1992): 589-618.
- Johnston, Mark (1997). "Human Concerns Without Superlative Selves." In Dancy (1997): 149-179.
- Keil, Frank C. (1989). *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: Bradford/MIT.
- Kolak, Daniel (1993a). "Finding Our Selves: Identification, Identity, and Multiple Personality." *Philosophical Psychology*, Vol. 6, No. 4 (1993): 363-386.
- Kolak, Daniel (1993b). "The Metaphysics and Metapsychology of Personal Identity: Why Thought Experiments Matter in Deciding Who We Are." *American Philosophical Quarterly*, Vol. 30, No. 1 (January 1993): 39-53.
- Kolak, Daniel and Raymond Martin (1987). "Personal Identity and Causality: Becoming Unglued." *American Philosophical Quarterly*, Vol. 24 (October 1987): 339-347.
- Kolak, Daniel and Raymond Martin, eds. (1991). *Self and Identity: Contemporary Philosophical Issues*. New York: Macmillan.
- Korsgaard, Christine (1989). "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy and Public Affairs*, Vol. 18, No. 2 (Spring 1989): 101-132.
- Kripke, Saul (1971). "Identity and Necessity." In Munitz (1971).
- Kripke, Saul (1972). *Naming and Necessity*. Cambridge: Harvard University Press.

- Lee, Win-Chiat (1990). "Personal Identity, the Temporality of Agency, and Moral Responsibility." *Auslegung*, Vol. 16, No. 1 (Winter 1990): 17-29.
- Lewis, David (1971). "Counterparts of Persons and their Bodies." *Journal of Philosophy*, Vol. 68: 203-211.
- Lewis, David (1976). "Survival and Identity." In Rorty (1976): 17-40. Reprinted with new pagination in Lewis (1983a): 55-72.
- Lewis, David (1983a). "Extrinsic Properties." *Philosophical Studies*, Vol. 44: 197-200.
- Lewis, David (1983b). *Philosophical Papers: Volume I*. New York and Oxford: Oxford University Press.
- Lewis, David (1983c). "Postscripts to 'Survival and Identity.'" In Lewis (1983a): 73-77.
- Lewis, David (1988). "Rearrangement of Particles: Reply to Lowe." *Analysis*, Vol. 48: 65-72.
- Lizza, John P. (1993). "Multiple Personality and Personal Identity Revisited." *British Journal for the Philosophy of Science*, Vol. 44, No. 2 (June 1993): 263-274.
- Lowe, E.J. (1983a). "On the Identity of Artifacts." *Journal of Philosophy*, Vol. 80 (April 1983): 220-231.
- Lowe, E.J. (1983b). "Instantiation, Identity and Constitution." *Philosophical Studies*, Vol. 44 (July 1983): 45-60.
- Lowe, E.J. (1988). "The Problems of Intrinsic Change: Rejoinder to Lewis." *Analysis*, Vol. 48 (March 1988): 72-77.
- Lowe, E.J. (1989). "What Is a Criterion of Identity?" *Philosophical Quarterly*, Vol. 39 (January 1989): 1-21.

- Lowe, E.J. (1991a). "Real Selves: Persons as a Substantial Kind." Reprinted in Cockburn (1991).
- Lowe, E.J. (1991b). "Substance and Selfhood." *Philosophy* (January 1991): 81-99.
- Lowe, E.J. (1995). "In Defence of the 'Standard Account'." *Analysis*, Vol. 55, No. 3 (July 1995): 171-178.
- Lowe, E.J. (1998). "Entity, Identity and Unity." *Erkenntnis*, Vol. 48, Nos. 2-3 (1998): 191-208.
- Lubcke, Poul (1993). "What Matters? On Parfit's Ideas of Personal Identity and Morality." *Danish Yearbook of Philosophy*, Vol. 28 (1993): 99-114.
- Lund, David H. (1994). *Perception, Mind and Personal Identity: A Critique of Materialism*. Lanham, Maryland: University Press of America.
- MacDonald, G.F., ed. (1979). *Perception and Identity: Essays Presented to A.J. Ayer and His Replies*. Ithaca, NY: Cornell University Press.
- MacKay, Donald (1987). "Divided Brains – Divided Minds?" In Blakemore and Greenfield (1987).
- Mackie, J. L. (1976). *Problems from Locke*. Oxford: Clarendon Press.
- Maddy, Penelope (1979). "Is the Importance of Identity Derivative?" *Philosophical Studies*, Vol. 35 (Fall 1979): 151-170.
- Madell, Geoffrey (1981). *The Identity of the Self*. Edinburgh: Edinburgh University Press.
- Madell, Geoffrey (1985). "Derek Parfit and Greta Garbo." *Analysis*, Vol. 45 (March 1985): 105-109.

Madell, Geoffrey (1989). *Personal Identity and the Mind-Body Problem in the Case for Dualism*. Charlottesville: University Press of Virginia.

Madell, Geoffrey (1991). "Personal Identity and the Idea of a Human Being." In Cockburn (1991): 127-142.

Margolis, J. (1988). "Minds, Selves and Persons." *Topoi*, Vol. 7 (March 1988): 31-45.

Marks, C. (1980). *Commissurotomy, Consciousness, and the Unity of Mind*. Cambridge, MA: MIT Press.

Martin, Raymond (1987). "Memory, Connecting, and What Matters in Survival." *Australasian Journal of Philosophy*, Vol. 65 (March 1987): 82-97.

Martin, Raymond (1988). "Identity's Crisis." *Philosophical Studies*, Vol. 53 (March 1988): 207-221.

Martin, Raymond (1990). "Identity, Transformation and What Matters in Survival." *Logos (USA)*, (1990): 57-70.

Martin, Raymond (1992). "Self-Interest and Survival." *American Philosophical Quarterly*, Vol. 29, No. 4 (October 1992): 165-184.

Martin, Raymond (1993a). "Having the Experience: The Next Best Thing to Being There." *Philosophical Studies*, Vol. 68, No. 3 (June 1993): 63-79.

Martin, Raymond (1993b). "Real Values: Why the Wilkes-Donagan Prohibition is Mistaken." *Metaphilosophy*, Vol. 24, No. 4 (October 1993): 400-406.

Martin, Raymond (1995). "Fission Rejuvenation." *Philosophical Studies*, Vol. 80, No. 1 (October 1995): 17-40.

Martin, Raymond (1998). *An Experiential Approach to What Matters in Survival*. Cambridge: Cambridge University Press.

Martin, Raymond, John Barresi, Alessandro Giovannelli, eds. (1998). "Fission Examples in the Eighteenth and Early Nineteenth Century Personal Identity Debate." *History of Philosophy Quarterly*, Vol. 15, No. 3 (July 1998): 323-348.

Matthews, Gareth B. (1977). "Surviving As." *Analysis*, Vol. 37 (January 1977): 53 -58.

Matthews, Steve (1998). "Personal Identity, Multiple Personality Disorder, and Moral Personhood." *Philosophical Psychology*, Vol. 11, No. 1 (March 1998): 67-88.

McCall, Catherine (1990). *Concepts of Person: An Analysis of Concepts of Person, Self, and Human Being*. Aldershot, UK: Avebury.

McDowell, John (1997). "Reductionism and the First Person." In Dancy (1997): 230-250.

Measor, Nicholas (1980). "On What Matters in Survival." *Mind*, Vol. 89 (July 1980): 406-411.

Merricks, Trenton (1997). "Fission and Personal Identity Over Time." *Philosophical Studies*, Vol. 88, No. 2 (November 1997): 163-186.

Merricks, Trenton (1998a). "Against the Doctrine of Microphysical Supervenience" *Mind*, Vol. 107 (January 1998): 59-71.

Merricks, Trenton (1998b). "There Are No Criteria of Identity Over Time." *Noûs*, Vol. 32, No. 1 (March 1998): 106-124.

Merricks, Trenton (1998c). "On Whether Being Conscious Is Intrinsic." *Mind*, Vol. 197, No. 428 (October 1998): 845-846.

Miller, Fred D. and Nicholas Smith, eds. (1989). *Thought Probes: Philosophy Through Science Fiction Literature*. Englewood Cliffs, NJ: Prentice Hall.

Mills, Eugene (1993). "Dividing without Reducing: Bodily Fission and Personal Identity." *Mind*, Vol. 102, No. 405 (January 1993): 37-51.

Miri, Mrinal (1973). "Memory and Personal Identity." *Mind*, Vol. 82 (January 1973): 1-21.

Montefiore, A., ed. (1973). *Philosophy and Personal Relations*. London: Routledge and Kegan Paul.

Moor, J. (1982). "Split Brains and Atomic Persons." *Philosophy of Science*, Vol. 49 (1982): 91-106.

Moran, Richard (1988). "Making Up Your Mind: Self-Interpretation and Self-Constitution." *Ratio* (New Series), Vol. I (December 1988): 135-151.

Munitz, Milton K., ed. (1971). *Identity and Individuation*. New York: New York University Press.

Nagel, Thomas (1971). "Brain Bisection and the Unity of Consciousness." *Synthese*, Vol. 22. Reprinted with new pagination in Perry (1975a): 227-245.

Nerlich, G. C. (1958). "Sameness, Difference, Continuity." *Analysis*, Vol. 18 (June 1958): 144-149.

Noonan, Harold (1978). "Locke on Personal Identity." *Philosophy*, Vol. 53 (July 1978): 343-351.

Noonan, Harold (1980). *Objects and Identity*. The Hague: Martinus Nijhoff Publishers.

Noonan, Harold (1981). "Wiggins' Second Thoughts on Identity." *Philosophical Quarterly*, Vol. 31 (July 1981): 260-268.

Noonan, Harold (1982). "Williams on 'The Self and the Future'." *Analysis*, Vol. 42 (June 1982): 158-163.

Noonan, Harold (1983). "Personal Identity and Bodily Continuity: A Further Note on 'The Self and the Future'." *Analysis*, Vol. 43, (March 1983): 98-104.

Noonan, Harold (1985a). "The Closest Continuer Theory of Identity." *Inquiry*, Vol. 28 (June 1985): 195-229.

Noonan, Harold (1985b). "Wiggins, Artefact Identity and 'Best Candidate' Theories." *Analysis*, Vol. 45 (January 1985): 4-8.

Noonan, Harold (1986). "Reply to Garrett's 'Noonan, 'Best Candidate' Theories and the Ship of Theseus'." *Analysis*, Vol. 46 (October 1986): 205-211.

Noonan, Harold (1988a). "Reply to Lowe on Ships and Structures." *Analysis*, Vol. 48 (October 1988): 221-223.

Noonan, Harold (1988b). "Substance, Identity and Time." *Aristotelian Society, Supplementary Vol.* 62 (1988): 79-100.

Noonan, Harold (1989). *Personal Identity*. London and New York: Routledge. Includes a comprehensive bibliography of related works, pg. 255-259.

Noonan, Harold (1993a). "Chisholm, Persons, and Identity." *Philosophical Studies*, Vol. 69, No. 1 (January 1993): 35-58.

Noonan, Harold, ed. (1993b). *Personal Identity*. Hampshire, UK: Ashgate Publishing Company.

- Nozick, Robert (1981). *Philosophical Explanations*. Cambridge, MA: Belknap Press.
- Oaklander, L. Nathan (1984). "Perry, Personal Identity and the 'Characteristic' Way." *Metaphilosophy*, Vol. 15 (January 1984): 35-44.
- Oaklander, L. Nathan (1987). "Parfit, Circularity, and the Unity of Consciousness." *Mind*, Vol. 96 (October 1987): 525-529.
- Oaklander, L. Nathan (1988). "Shoemaker on the Duplication Argument: Survival and What Matters." *Australasian Journal of Philosophy*, Vol. 66 (June 1988): 234-239.
- Odegard, Douglas (1969). "Personal and Bodily Identity." *American Philosophical Quarterly*, Vol. 19 (January 1969): 69-71.
- Odegard, Douglas (1972). "Identity Through Time." *American Philosophical Quarterly*, Vol. 9 (January 1972): 29-38.
- Oderberg, David S. (1989a). "Johnston on Human Beings." *Journal of Philosophy*, Vol. 86, No. 3 (March 1989): 137-141.
- Oderberg, David S. (1989b). "Reply to Sprigge on Personal and Impersonal Identity." *Mind*, Vol. 98 (January 1989), 129-133.
- Oderberg, David S. (1993). *The Metaphysics of Identity over Time*. New York: St. Martin's Press.
- Oderberg, David S. (1996). "Coincidence under a Sortal." *Philosophical Review*, Vol. 105: 145-171.
- Olson, Eric (1994). "Is Psychology Relevant to Personal Identity?" *Australasian Journal of Philosophy*, Vol. 72 (1994): 173-186.

- Olson, Eric (1995). "Human People or Human Animals?" *Philosophical Studies*, Vol. 80, No. 2 (November 1995): 159-181.
- Olson, Eric (1997a). *The Human Animal: Personal Identity Without Psychology*. Oxford: Oxford University Press.
- Olson, Eric (1997b). "Relativism and Persistence." *Philosophical Studies*, Vol. 88, No. 2 (November 1997): 141-162.
- Olson, Eric (1997c). "Was I Ever A Fetus?" *Philosophy and Phenomenological Research*, Vol. 57, No. 1 (March 1997): 95-110.
- Olson, Eric (1998). "Human Atoms." *Australasian Journal of Philosophy*, Vol. 76, No. 3 (September 1998): 396-406.
- Parfit, Derek (1971a). "On the Importance of Self-Identity." *Journal of Philosophy*, Vol. 68 (October 1971): 683-690.
- Parfit, Derek (1971b). "Personal Identity." *Philosophical Review*, Vol. 80, No. 1 (January, 1971). Reprinted with new pagination in Perry (1975a): 199-223.
- Parfit, Derek (1973). "Later Selves and Moral Principles." In Montefiore (1973).
- Parfit, Derek (1976). "Lewis, Perry and What Matters." In Rorty (1976): 91-107.
- Parfit, Derek (1982). "Personal Identity and Rationality." *Synthese*, Vol. 53 (November 1982): 227-24.
- Parfit, Derek (1984/1987). *Reasons and Persons*. Oxford: Clarendon Press.
- Parfit, Derek (1986). "Comments." *Ethics*, Vol. 96 (July 1986): 832-872.
- Parfit, Derek (1987). "Divided Minds and the Nature of Persons." In Blakemore and Greenfield (1987).

Parfit, Derek (1992) "Who Do You Think You Are?" *Times Higher Education Supplement*, Vol. 11 (December 1992): 19-20.

Parfit, Derek (1993). "The Indeterminacy of Identity: A Reply to Brueckner." *Philosophical Studies*, Vol. 70, No. 1 (April 1993): 23-33.

Parfit, Derek (1994). "Personal Identity." Unpublished manuscript.

Parfit, Derek (1995). "The Unimportance of Identity." In Harris (1995b): 13-47.

Peacocke, A and G. Gillett, eds. (1987b). *Persons and Personality: A Contemporary Inquiry*. Oxford: Basil Blackwell.

Penelhum, Terence (1955). "Hume on Personal Identity." *Philosophical Review*, Vol. 64 (October 1955): 571-589.

Penelhum, Terence (1957). "Is It Possible That One and the Same Individual Object Should Cease to Exist and Later On, Start to Exist Again?" *Analysis*, Vol. 17 (June 1957): 123-124.

Penelhum, Terence (1959). "Personal Identity, Memory and Survival." *Journal of Philosophy*, (October, 1989): 882-902.

Penelhum, Terence (1970). *Survival and Disembodied Existence*. London: Routledge and Kegan Paul; New York: Humanities Press.

Penelhum, Terence (1971). "The Importance of Self-Identity." *Journal of Philosophy*, Vol. 68 (October 1971): 667-678.

Penelhum, Terrence (1976). "Self-Identity and Self-Regard." In Rorty (1976): 253-280.

Perry, John (1972). "Can the Self Divide?" *Journal of Philosophy*, Vol. 69, No. 16 (September 7, 1972): 463-488.

Perry, John, ed. (1975a). *Personal Identity*. Berkeley and Los Angeles: University of California Press.

Perry, John (1975b). "Personal Identity, Memory and the Problem of Circularity." In Perry (1975a): 135-155.

Perry, John (1975c). "The Problem of Personal Identity." In Perry (1975a): 3-30.

Perry, John (1976). "The Importance of Being Identical." In Rorty (1976): 67-90.

Perry, John (1978). *A Dialogue on Personal Identity and Immortality*. Indianapolis: Hackett.

Perry, John (1983). "Personal Identity and the Concept of a Person." In Guttorm Floistad, ed., *A New Survey*, 11-43.

Persson, Ingmar (1992). "The Indeterminacy and Insignificance of Personal Identity." *Inquiry*, Vol. 35, No. 2 (June 1992): 249-269.

Persson, Ingmar (1993). "Critical Study of Peter van Inwagen's Material Beings." *Noûs*, Vol. 27, No. 4: 512-518.

Persson, Ingmar (1997a). "Genetic Therapy, Person-Regarding Reasons and the Determination of Identity: A Reply to Robert Elliot." *Bioethics*, Vol. 11, No. 2 (April 1997): 161-169.

Persson, Ingmar (1997b). "The Involvement of Our Identity in Experiential Memory." *Canadian Journal of Philosophy*, Vol. 27, No. 4 (December 1997).

- Peterson, J. (1985). "Persons and the Problem of Interaction." *Modern Schoolman*, Vol. 62 (1985): 131-138.
- Pogue, J. E. (1993). "Identity, Survival, and the Reasonableness of Replication." *Southern Journal of Philosophy*, Vol. 31, No. 1 (Spring 1993): 45-70.
- Puccetti, Roland (1969). "Brain Transplantation and Personal Identity." *Analysis*, Vol. 29, No. 3 (New Series): 65-77.
- Puccetti, Roland (1973a). "Brain Bisection and Personal Identity." *British Journal for the Philosophy of Science*, Vol. 24 (1973): 339-355.
- Puccetti, Roland (1973b). "Multiple Identity." *Personalist*, Vol. 54 (1973): 203-213.
- Puccetti, Roland (1975). "The Mute Self: A Reaction to DeWitt's Alternative Account of the Split-Brain Data." *British Journal for the Philosophy of Science*, Vol. 27 (1975): 65-73.
- Puccetti, Roland (1981). "The Case for Mental Duality: Evidence from the Split-Brain Data and Other Considerations." *Behavioral and Brain Sciences*, Vol. 4 (1981): 93-123.
- Puccetti, Roland (1989). "Two Brains, Two Minds." *British Journal for the Philosophy of Science*, Vol. 40 (1989): 137-144.
- Puccetti, Roland (1993a). "Dennett on the Split-Brain." *Psychology*, Vol. 4, No. 52.
- Puccetti, Roland (1993b). "Mind With a Double Brain." *British Journal for the Philosophy of Science*, Vol. 44 (1993): 675-692.

- Quine, W.V. (1972). "Review of Identity and Individuation." *Journal of Philosophy*, Vol. 69, No. 16 (September 7, 1972): 488-497.
- Quinton, Anthony (1962/1975). "The Soul." *Journal of Philosophy*, Vol. 59, No. 15 (July 1962). Reprinted with new pagination in Perry (1975a): 53-72.
- Radden, J. (1996). *Divided Minds and Successive Selves: Ethical Issues in Disorders of Identity and Personality*. Cambridge, MA: MIT Press.
- Rea, Michael C. (1995). "The Problem of Material Constitution." *Philosophical Review*, Vol. 104, No. 4 (October 1995): 525-553.
- Rea, Michael C., ed. (1997) *Material Constitution: A Reader*. Savage, Maryland: Rowman & Littlefield.
- Rea, Michael C. (1998). "Sameness Without Identity: An Aristotelian Solution to the Problem of Material Constitution." *Ratio*, Vol. 11, No. 3 (December 1998): 316-328.
- Reid, Mark (1997). "Narrative and Fission: A Review Essay of Marya Schechtman's 'The Constitution of Selves'." *Philosophical Psychology*, Vol. 10, No. 2 (June 1997): 211-219.
- Reid, Thomas (1785/1975). "Of Identity." Reprinted in Perry (1975a): 107-112.
- Reid, Thomas (1785/1975). "Of Mr. Locke's Account of Our Personal Identity." Reprinted in Perry (1975a): 113-118.
- Reuscher, John A. (1981). *Essays on the Metaphysical Foundation of Personal Identity*. Washington, D.C.: University Press of America.

- Rey, Georges (1976). "Survival." In Rorty (1976): 41-66.
- Rieber, Steven (1998). "The Concept of Personal Identity." *Philosophy and Phenomenological Research*, Vol. 58, No. 3 (September 1998): 581-594.
- Robert, M. (1983). "Lewis's Theory of Personal Identity." *Australasian Journal of Philosophy*, Vol. 61 (1983): 58-67.
- Robinson, Denis (1976). "What Sort of Persons are Hemispheres? Another Look at 'Split-Brain' Man." *British Journal for the Philosophy of Science*, Vol. 27 (1976): 73-78.
- Robinson, Denis (1982). "Reidentifying Matter." *Philosophical Review*, Vol. 91 (1982): 317-341.
- Robinson, Denis (1985). "Can Amoeba Divide without Multiplying?" *Australasian Journal of Philosophy*, Vol. 63: 299-319.
- Robinson, John (1988). "Personal Identity and Survival." *Journal of Philosophy*, Vol. 85, No. 6 (June 1988): 319-328.
- Rorty, Amélie Oksenberg, ed. (1976). *The Identities of Persons*. Berkeley: University of California Press. Includes a comprehensive bibliography of related works published prior to 1971 on pg. 325-333.
- Rosenberg, Jay F. (1993). "Comments on Peter van Inwagen's Material Beings." *Philosophy and Phenomenological Research*, Vol. 53, No. 3 (September 1993): 701-708.
- Rovane, Carol (1990). "Branching Self-Consciousness." *Philosophical Review*, Vol. 99, No. 3 (July 1990): 355-395.

- Rovane, Carol (1993). "Self-Reference: The Radicalization of Locke." *Journal of Philosophy*, Vol. 90, No. 2 (February 1993): 73-97.
- Rovane, Carol (1994). "Critical Notice of Peter Unger's Identity, Consciousness and Value." *Canadian Journal of Philosophy*, Vol. 24, No. 1 (March 1994): 119-133.
- Rovane, Carol (1998). *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton, NJ: Princeton University Press.
- Salmon, Nathan (1982). *Reference and Essence*. Oxford: Basil Blackwell.
- Scaltsas, Theodore (1980). "The Ship of Theseus." *Analysis*, Vol. 40, No. 3 (June 1980): 152-157.
- Schechtman, Marya (1990). "Personhood and Personal Identity." *Journal of Philosophy*, Vol. 77, No. 2 (February 1990): 71-92.
- Schechtman, Marya (1994). "The Same and the Same: Two Views of Psychological Continuity." *American Philosophical Quarterly*, Vol. 31, No. 3 (July 1994): 199-212.
- Schechtman, Marya (1996). *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
- Sedley, David (1982). "The Stoic Criterion of Identity." *Phronesis*, Vol. 22 (1982): 255-275.
- Shaffer, J. (1977). "Personal Identity: The Implications of Brain Bisection and Brain Transplants." *Journal of Medicine and Philosophy*, Vol. 2 (1977): 147-161.
- Shalom, Albert (1985). *The Body-Mind Conceptual Framework and the Problem of Personal Identity*. Atlantic Highlands: Humanities Press.

- Shoemaker, David (1996). "Theoretical Persons and Practical Agents." *Philosophy and Public Affairs*, Vol. 25, No. 4 (Fall 1996): 318-332.
- Shoemaker, Sydney (1959/1975). "Personal Identity and Memory." *Journal of Philosophy*, Vol. LVI, No. 22 (October 22, 1959). Reprinted with new pagination in Perry (1975a): 119-134.
- Shoemaker, Sydney (1963). *Self-Knowledge and Self-Identity*. Ithaca: Cornell University Press.
- Shoemaker, Sydney (1970/1984). "Persons and Their Pasts." *American Philosophical Quarterly*, Vol. 7, No. 4 (October 1970): 269-285. Reprinted with new pagination in Shoemaker (1984): 19-48.
- Shoemaker, Sydney (1971). "Wiggins on Identity." In Munitz (1971): 103-118.
- Shoemaker, Sydney (1976). "Embodiment and Behavior." In Rorty (1976): 109-137.
- Shoemaker, Sydney (1984). *Identity, Cause and Mind*. Cambridge: Cambridge University Press.
- Shoemaker, Sydney (1985). "Critical Notice of Reasons and Persons." *Mind*, Vol. 94 (1985): 443-453
- Shoemaker, Sydney (1992). "Unger's Psychological Continuity Theory." *Philosophy and Phenomenological Research*, Vol. LII, No. 1 (March 1992): 139-143.
- Shoemaker, Sydney (1994). "The First-Person Perspective." *Proceedings of the APA*, Vol. 68, No. 2 (November 1994): 7-22.
- Shoemaker, Sydney (1997a). "Parfit on Identity." In Dancy (1997): 135-148.
- Shoemaker, Sydney (1997b). "Self and Substance." In Tomberlin (1997).
- Shoemaker, Sydney and Richard Swinburne (1984). *Personal Identity: Great Debates in Philosophy*. Oxford: Basil Blackwell.
- Shorter, J.M. (1962). "More About Bodily Continuity and Personal Identity." *Analysis*, Vol. 22 (1962): 79-85.
- Sider, Theodore. "Recent Work in Identity Over Time." *Philosophical Books*. Vol. 41 (2000): 81-89.
- Siderits, M. (1988). "Ehring on Parfit's Relation R." *Analysis*, Vol. 48 (January 1988): 29-32.
- Simons, Peter (1987). *Parts: A Study in Ontology*. Oxford: Oxford University Press.
- Smart, Brian (1972). "How to Reidentify the Ship of Theseus." *Analysis*, Vol. 32, No. 5 (April 1972): 145-148.
- Smart, Brian (1973a). "Personal Identity in an Organized Parcel." *Philosophical Studies*, Vol. 24 (November 1973): 420-423.
- Smart, Brian (1973b). "The Ship of Theseus, the Parthenon, and Disassembled Objects." *Analysis*, Vol. 34, No. 1 (October 1973): 24-27.
- Smart, Brian (1974). "Persons and Selves." *Philosophical Studies*, Vol. 26 (December 1974): 331-336.
- Smart, Brian (1975). "The Statues and a Problem of Identity." *Ratio*, Vol. 17 (December 1975): 229-236.
- Smart, Brian (1976). "Synchronous and Diachronous Selves." *Canadian Journal of Philosophy*, Vol. 6 (1976): 13-33.

Snowdon, Paul F. (1990). "Persons, Animals and Ourselves." In Gill (1990).

Snowdon, Paul F. (1991). "Personal Identity and Brain Transplants." In Cockburn (1991):109-126.

Sorensen, Roy (1992). *Thought Experiments*. New York and Oxford: Oxford University Press.

Sosa, Ernest (1990). "Surviving Matters." *Noûs*, Vol. 24, No. 2 (April 1990): 305-330.

Sosa, Ernest (1997). "Persons and Other Beings." In Tomberlin (1989): 155-187.

Sousa, Ronald de (1976). "Rational Homunculi." In Rorty (1976): 217-238.

Sperry, Roger W. (1984). "Consciousness, Personal Identity, and the Divided Brain." *Neuropsychologia*, Vol. 22 (1984): 611-673.

Sprigge, T.L.S. (1988). "Personal and Impersonal Identity." *Mind*, Vol. 98 (October 1988): 29-49.

Sprigge, T.L.S. (1989). "Personal and Impersonal Identity: A Reply to Oderberg." *Mind*, Vol. 98 (1989): 605-607.

Stone, J. (1988). "Parfit and the Buddha: Why There are No People." *Philosophy and Phenomenological Research*, Vol. 48 (March 1988): 519-532.

Storl, Heidi (1992). "The Problematic Nature of Parfitian Persons." *Personalist Forum*, Vol. 8, No. 1 (Supplement) (Spring 1992): 123-131.

Strawson, Galen (1999a). "The Sense of the Self." In Crabbe (1999).

Strawson, Galen (1999b). "The Self." In Gallagher and Shear (1999): 1-24.

Strawson, P.F. (1958). "Persons." *Minnesota Studies in Philosophy of Science*, Vol. 2 (1958): 330-353.

Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.

Strawson, P.F. (1966). "Self, Mind, and Body." *Common Factor*, Vol. 4 (1966): 5-13.

Strawson, P. F. (1992). "Comments on Some Aspects of Peter Unger's Identity Consciousness and Value." *Philosophy and Phenomenological Research*, Vol. 52, No. 1 (March 1992): 145-148.

Swinburne, Richard (1992). Discussion of Peter Unger's Identity Consciousness and Value." *Philosophy and Phenomenological Research*, Vol. 52, No. 1 (March 1992): 149-152.

Swoyer, Chris (1982). "Hume and Three Views of the Self." *Hume Studies* 8: 43-61.

Swoyer, Chris (1984). "Causation and Identity." In French et al, eds. (1984): 593-622.

Taylor, Charles (1976). "Responsibility for Self." In Rorty (1976): 281-300.

Taylor, Charles (1977). "Self-Interpreting Animals." Reprinted in Taylor (1985): 45-76.

Taylor, Charles (1981). "The Concept of a Person." Reprinted in Taylor (1985): 97-114.

Taylor, Charles (1985). *Human Agency and Language: Philosophical Papers*, Volume 1. Cambridge: Cambridge University Press.

Thiel, Udo (1998). "Personal Identity." In Garber, Daniel ed. *The Cambridge History of Seventeenth-Century Philosophy: Two Volumes.*

Thomson, Judith Jarvis (1983). "Parthood and Identity Across Time." *Journal of Philosophy* 80: 201-220.

Thomson, Judith Jarvis, ed. (1987a). *On Being and Saying: Essays in Honor of Richard Cartwright.* Cambridge, MA: MIT Press.

Thomson, Judith Jarvis, (1987b). "Ruminations on an Account of Personal Identity." In Thomson (1987a): 215-240.

Thomson, Judith Jarvis (1997). "People and Their Bodies." In Dancy (1997): 202-229.

Thomson, Judith Jarvis (1998). "The Statue and the Clay." *Noûs* 32: 149-173.

Tomberlin, James E., ed. (1997). *Mind, Causation and World.* Boston, MA: Blackwell.

Trupp, Andreas (1987). *Why We Are What We Think We Are: A New Approach to the Nature of Personal Identity and of Time.* Frankfurt am Main; New York: P. Lang.

Tye, Michael (1992). "Review of Peter van Inwagen's Material Beings." *Philosophical Review*, Vol. 101, No. 4 (October 1992): 881-884.

Unger, Peter (1979a). "I Do Not Exist." In MacDonald (1979).

Unger, Peter (1979b). "Why There Are No People." *Midwest Studies in Philosophy*, Vol. 4 (1979): 177-222.

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

Unger, Peter (1992). "Précis of Identity Consciousness and Value" and "Reply to Reviewers." *Philosophy and Phenomenological Research*, Vol. 52, No. 1 (March 1992): 133-137, 159-176.

Van Inwagen, Peter (1980a). "Philosophers and the Words 'Human Body'." In Van Inwagen (1980b): 283-300.

Van Inwagen, Peter, ed. (1980b). *Time and Cause: Essays Presented to Richard Taylor.* Dordrecht: Reidel.

Van Inwagen, Peter (1990). *Material Beings.* Ithaca, NY: Cornell University Press.

Van Inwagen, Peter (1993a). "Critical Study of Peter Unger's Identity Consciousness and Value." *Noûs*, Vol. 27, No. 2: 373-379

Van Inwagen, Peter (1993b). "Précis of Material Beings" and "Reply to Reviewers" *Philosophy and Phenomenological Research*, Vol. 53, No. 3 (September 1993): 683-691, 709-719.

Van Inwagen, Peter (1997). "Materialism and the Psychological-Continuity Account of Personal Identity." In Tomberlin (1997).

Velleman, J. David (1996). "Self to Self." *Philosophical Review*, Vol. 105, No. 1 (1996): 39-76.

Vesey, Godfrey (1974). *Personal Identity: A Philosophical Analysis.* Ithaca, NY: Cornell University Press.

White, Stephen (1989). "Metapsychological Relativism and the Self." *Journal of Philosophy*, Vol. 86 (1989): 298-322.

White, Stephen (1995). "The Desire to Survive." *Philosophy and Phenomenological Research*, Vol. 52, No. 1 (March 1992): 153-158.

- Whiting, Jennifer (1986). "Friends and Future Selves." *Philosophical Review*, Vol. 95, No. 4 (October 1986): 547-580.
- Whiting, Jennifer (1991). "Impersonal Friends." *The Monist*, Vol. 74, (January 1986): 3-29.
- Wiggins, David (1967). *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.
- Wiggins, David (1968). "On Being in the Same Place at the Same Time." *Philosophical Review*, Vol. 77 (1968): 90-95.
- Wiggins, David (1976). "Locke, Butler and the Stream of Consciousness: And Men as Natural Kind." In Rorty (1976): 139-173.
- Wiggins, David (1979). "The Concern to Survive." *Midwest Studies in Philosophy*, Vol. 4 (1979): 417-422.
- Wiggins, David (1980). *Sameness and Substance*. Cambridge, MA: Harvard University Press.
- Wiggins, David (1987). "The Person as Object of Science, as Subject of Experience, and as Locus of Value." In Peacocke and Gillett (1987).
- Wilkes, Kathleen V. (1988). *Real People: Personal Identity without Thought Experiments*. Oxford: Clarendon Press.
- Wilkes, Kathleen V. (1991a). "How Many Selves Make Me?" *Philosophy*, Vol. 66 (Supplement 1991): 235-243.
- Wilkes, Kathleen V. (1991b). "Multiple Personality and Personal Identity." *British Journal for the Philosophy of Science*, Vol. 32 (1991): 331-348.

- Williams, Bernard (1956/1973). "Personal Identity and Individuation." Reprinted (with new pagination) in Williams (1973): 1-18.
- Williams, Bernard (1960/1973). "Bodily Continuity and Personal Identity." Reprinted (with new pagination) in Williams (1973): 19-25.
- Williams, Bernard (1966/1973). "Imagination and the Self." Reprinted (with new pagination) in Williams (1973): 26-45.
- Williams, Bernard (1970a/1973). "Are Persons Bodies?" Reprinted (with new pagination) in Williams (1973): 64-81.
- Williams, Bernard (1970b/1973). "The Self and the Future." Reprinted (with new pagination) in Williams (1973): 46-63. Also in Perry (1975a): 179-198.
- Williams, Bernard (1973). *Problems of the Self*. Cambridge: Cambridge University Press.
- Williams, Bernard (1976). "Persons, Character and Morality." In Rorty (1976): 197-216.
- Williams, Bernard (1995). "Identity and Identities." In Harris (1995b): 1-13.
- Williamson, Tim (1990). *Identity and Discrimination*. New York: Blackwell.
- Wolf, Susan (1986). "Self-Interest and Interest in Selves." *Ethics*, 96 (July 1986): 704-720.
- Zemach, Eddy M. (1987). "Looking Out For Number One." *Philosophy and Phenomenological Research*, Vol. 48 (December 1987): 209-233.

Zuboff, Arnold (1978). "Moment Universals and Personal Identity." *Proceedings of the Aristotelian Society*, Vol. 78 (1977-78): 141-155.

Zuboff, Arnold (1990). "One Self: The Logic of Experience." *Inquiry*, Vol. 33, No. 1 (March 1990): 39-68.

6.4 BIBLIOGRAPHY ON THOUGHT EXPERIMENT AND EXPERIMENT

(Prepared by David Bzdak)

Achinstein, Peter, and Owen Hannaway, eds. (1985).

Observation, Experiment, and Hypothesis in Modern Physical Science. Cambridge, MA: MIT Press.

Akhutin, A.V. (1982). "Galileo's Experiment." *Scientia*, Vol. 117 (1982): 287-312.

Anapolitanos, D. A. (1991). "Thought Experiments and Conceivability Conditions in Mathematics." In Horowitz and Massey (1991): 87-99.

Baillie, James (1991). Split Brains and Single Minds." *Journal of Philosophical Research*, Vol. 16 (1991): 11-18.

Baillie, James (1993). "Recent Work on Personal Identity." *Philosophical Books*, Vol. 34, No. 4 (October, 1993): 193-206.

Bealer, George (1998). "The Autonomy of Philosophy." In DePaul and Ramsey (1998): 201-240.

Beck, Simon (1992). "Should We Tolerate People Who Split?" *The Southern Journal of Philosophy*, Vol. 30, No. 1 (Spring 1992): 1-17.

Beck, Simon (1998). "Back to the Self and the Future." *South African Journal of Philosophy*, Vol. 17, No. 3 (August 1998): 211-225.

Bishop, Michael A. (1998). "An Epistemological Role for Thought Experiments." In Shanks (1998).

Born, Max (1956). *Experiment and Theory in Physics*. New York: Dover.

- Bourgeois, Warren (1995). *Persons, What Philosophers Say About You*. Waterloo, ON: Wilfrid Laurier University Press.
- Brooks, D.M.H. (1994). "The Method of Thought Experiment." *Metaphilosophy*, Vol. 25, No. 1 (January 1994): 71-83.
- Brown, James Robert (1986). "Thought Experiments Since the Scientific Revolution." *International Studies in the Philosophy of Science*, Vol. 1 (September 1986): 1-15.
- Brown, James Robert (1991a). *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. New York and London: Routledge.
- Brown, James Robert (1991b). "Thought Experiments: A Platonic Account." In Horowitz and Massey (1991): 119-128.
- Brown, James Robert (1993a). "Author's Response (to Norton [1993])." *Metascience* (new series), Vol. 3 (1993): 38-40.
- Brown, James Robert (1993b). "Why Empiricism Won't Work." *Proceedings of the Philosophy of Science Association* 2, 271-279.
- Brown, James Robert (1995). "Critical Notice of Roy Sorensen 'Thought Experiments'." *Canadian Journal of Philosophy*, Vol. 25, No. 1 (March 1995): 135-142.
- Broyles, James (1974). "An Observation on Wittgenstein's Use of Fantasy." *Metaphilosophy*, Vol. 5 (1974): 291-297.
- Broyles, James (1986). "Wittgenstein On Personal Identity." *Philosophical Investigations*, Vol. 9 (January 1986): 56-65.
- Bunzl, Martin (1995). "Bunzl on Sorensen's Thought Experiments." *Informal Logic*, Vol. 17, No. 3 (Fall 1995): 388-393.
- Bunzl, Martin (1996). "The Logic of Thought Experiments." *Synthese*, Vol. 106, No. 2 (Fall 1996): 227-240.
- Cargile, James (1987). "Definitions and Counterexamples." *Philosophy*, Vol. 62: 179-193.
- Carrier, Martin (1993). Review of Horowitz and Massey 1991. *Erkenntnis*, Vol. 39 (1993): 413-419.
- Cartwright, Helen M. (1993). "On Two Arguments for the Indeterminacy of Personal Identity." *Synthese*, Vol. 95, No. 2 (May 1993): 241-273.
- Chandra, Suresh (1977). *Identity and Thought Experiment*. Simla: Indian Institute for Advanced Study.
- Cohen, Robert S., ed. (1980). *Experiment, Theory, Practice: Articles and Addresses*. Dordrecht: Reidel.
- Cole, David (1984). "Thought and Thought Experiments." *Philosophical Studies*, Vol. 45 (1984): 431- 444.
- Cummins, Robert (1998). "Reflections on Reflective Equilibrium." In DePaul and Ramsey (1998): 113-128.
- Damerow, Peter et al. (1992). *Exploring the Limits of Pre-Classical Mechanics: A Study of Conceptual Development in Early Modern Science: Free Fall and Compounded Motion in the Work of Descartes, Galileo and Beeckman*. New York: Springer Verlag.
- Dancy, Jonathan (1985a). *An Introduction to Contemporary Epistemology*. Oxford: Basil Blackwell.

- Dancy, Jonathan (1985b). "The Role of Imaginary Cases in Ethics." *Pacific Philosophical Quarterly*, Vol. 66 (January-April 1985): 141-153.
- Dancy, Jonathan (1997). *Reading Parfit*. Oxford: Basil Blackwell.
- Danly, John R. (1985). "An Examination of the Fundamental Assumption of Hypothetical Process Arguments." *Philosophical Studies*, Vol. 48 (1985): 83 -89.
- Dear, Peter, ed. (1991). *The Literary Structure of Scientific Argument: Historical Studies*. Philadelphia: University of Pennsylvania Press.
- Dennett, Daniel (1978). *Brainstorms*. Cambridge, MA: Bradford Books
- Dennett, Daniel (1991). *Consciousness Explained*. Boston: Little Brown.
- DePaul, Michael R. and William Ramsey, eds. (1998). *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.
- Dijksterhuis, E. J. (1961, repr. 1986). *The Mechanization of the World Picture: Pythagoras to Newton*. Trans. C. Dikshoorn. Princeton, NJ: Princeton University Press.
- Dorries, Matthias (1994). "Balances, Spectroscopes, and the Reflexive Nature of Experiment." *Studies in History and Philosophy of Science*, Vol. 25, No. 1 (February 1994): 1-36.
- Drake, Stillman (1989). *History of Free Fall: Aristotle to Galileo*. Toronto: Wall & Thompson.

- Earman, John et al eds. (1993). *Philosophical Problems of the Internal and External Worlds: Essays on the Philosophy of Adolf Grünbaum*. Pittsburgh: University of Pittsburgh Press.
- Elkana, Yehuda, ed. (1974). *The Interaction Between Science and Philosophy*. Highlands, NJ: Humanities Press.
- Feldman, Richard (1995). "Feldman on Sorensen's Thought Experiments." *Informal Logic*, Vol. 17, No. 3 (Fall 1995): 394-398.
- Fiala, Silvo (1951). "The Experiment and Its Role in the Theory of Knowledge." *Philosophy of Science*, Vol. 18 (July 1951): 253-258.
- Fine, Kit (1975). "Critical Notice of Counterfactuals." *Mind*, Vol. 84 (1975): 451-458.
- Fodor, Jerry A. (1971). "On Knowing What We Would Say." In Rosenberg and Travis (1971): 198-212.
- Foot, Philippa (1978a). *Virtues and Vices*. Oxford: Blackwell.
- Foot, Philippa (1978b). "The Problem of Abortion and the Doctrine of Double Effect." In Foot (1978a).
- Franklin, Allan (1981). "What Makes a 'Good' Experiment?" *British Journal for the Philosophy of Science*, Vol. 32 (1981): 367-374.
- Franklin, Allan (1986). *The Neglect of Experiment*. Cambridge: Cambridge University Press.
- Gale, Richard (1991). "On Some Pernicious Thought Experiments." In Horowitz and Massey (1991): 297-304.
- Gallagher, Shaun and Jonathan Shear, eds. (1999). *Models of the Self*.

Gendler, Tamar Szabó (1998a). "Exceptional Persons: On the Limits of Imaginary Cases." *Journal of Consciousness Studies*, Vol., 5, Nos. 5-6 (1998): 592-610. Reprinted in Gallagher and Shear (1999): 447-466.

Gendler, Tamar Szabó (1998b). "Galileo and the Indispensability of Scientific Thought Experiment." *British Journal for the Philosophy of Science*, Vol. 49, No. 3 (September 1998): 397-424.

Gentner, Dedre and Albert L. Stevens, eds. (1983). *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Giere, R. ed. (1992). *Cognitive Models of Science*. Minneapolis: University of Minnesota Press.

Gillett, Grant (1986a). "Brain Bissection and Personal Identity." *Mind*, Vol. 95 (April 1986): 224-229.

Goldman, Alvin (1989). "Psychology and Philosophical Analysis." *Proceedings of the Aristotelian Society*, 89 (1998-99): 195-209.

Goldman Alvin and Joel Pust (1998). "Philosophical Theory and Intuitive Evidence." In DePaul and Ramsey (1998): 179-200.

Gomila, Antoni (1991). "What Is a Thought Experiment?" *Metaphilosophy*, (January-April 1991): 84-92.

Gooding, David (1990). *Science and Philosophy: Experiment and the Making of Meaning*. Dordrecht: Kluwer.

Gooding, David C. (1992). "The Cognitive Turn, or, Why Do Thought Experiments Work?" In Giere (1992): 45-76.

Gooding, David C. (1993). "What is Experimental about Thought Experiments?" *Proceedings of the Philosophy of Science Association*, Vol. 2 (1993): 280-290.

Gooding, David C. (1994). "Imaginary Science." *British Journal for the Philosophy of Science*, Vol. 45, No. 4 (December 1994): 1029-1045.

Gooding, David C., Pinch, T. and Schaffer, S. (1989). *The Uses of Experiment: Studies in the Natural Sciences*. Cambridge: Cambridge University Press.

Gopnik, Alison and Eric Schwitzgebel (1998). "Whose Concepts are They, Anyway? The Role of Philosophical Intuition in Empirical Psychology." In DePaul and Ramsey (1998): 75-94.

Hacking, Ian (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hacking, Ian (1993). "Do Thought Experiments Have a Life of Their Own?" *Proceedings of the Philosophy of Science Association*, Vol. 2 (1993): 302-308.

Häggqvist, Sören (1996). *Thought Experiments in Philosophy*. Stockholm: Almqvist & Wiksell International.

Haksar, Vinit (1991). *Indivisible Selves and Moral Practices*. Edinburgh: Edinburgh University Press.

Hanson, Norwood Russell (1958). *Patterns of Discovery*. Cambridge: Cambridge University Press.

Hansson, Sven Ove (1994). "Philosophical Craftsmanship." *Metaphilosophy*, Vol. 25, No. 4 (October 1994): 316-325.

Harman, Gilbert (1986). "Moral Explanations of Natural Facts: Can Moral Claims Be Tested Against Reality?" *Southern Journal of Philosophy* 24, Supplement.

Harré, Rom (1998). "Recovering the Experiment." *Philosophy*, Vol. 73, No. 285 (July 1998): 353-377.

- Harris, Henry (1995). "An Experimentalist Looks at Identity." In Harris (1995b): 47-65.
- Hart, Ivor B. (1923). *Makers of Science: Mathematics, Physics, Astronomy*. London: Oxford University Press.
- Hertzberg, Lars (1991). "Imagination and the Sense of Identity." In Cockburn (1991): 143-156.
- Hesse, Mary (1966). *Models and Analogies in Science*. Notre Dame: University of Notre Dame Press.
- Hiebert, Erwin (1974). "Mach's Conception of Thought Experiments in the Natural Sciences." In Elkana (1974).
- Hill, Christopher S. (1997). "Imaginability, Conceivability, Possibility and the Mind- Body Problem." *Philosophical Studies*, Vol. 87, No. 1 (July 1997): 61-85.
- Hintikka, Jaakko (1999). "The Emperor's New Intuitions." *Journal of Philosophy*, Vol. 96, No. 3. (1999): 127-147.
- Hirsch, Eli (1982). *The Concept of Identity*. New York: Oxford University Press.
- Holton, Gerald (1988). *Thematic Origins of Scientific Thought*. Cambridge: Harvard University Press.
- Horgan, Terence (1993). "On What There Isn't." *Philosophy and Phenomenological Research*, Vol. 53, No. 3 (September 1993): 693-700.
- Horowitz, Tamara and Gerald Massey, eds. (1991). *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield Publishers.
- Horowitz, Tamara (1998). "Philosophical Intuitions and Psychological Theory." *Ethics*, Vol. 108, No. 2 (January 1998): 367-385. A version of this paper also appears in DePaul and Ramsey (1998), pp. 143-160.

- Huang, Yih-mei (1995). "Personal Identity and Thought Experiments." *Philosophical Review (Taiwan)*, Vol. 18 (January 1995): 93-110.
- Hull, David L. (1989a). "A Function for Actual Examples in Philosophy of Science." In Hull (1989b).
- Hull, David L., ed. (1989b). *What the Philosophy of Biology Is*. Dordrecht: Kluwer Academic.
- Humphreys, Paul (1993). "Seven Theses on Thought Experiments." In Earman et al (1993): 205-227.
- Irvine, Andrew D. (1991). "On the Nature of Thought Experiments in Scientific Reasoning" in Horowitz and Massey (1991): 149-165.
- Jackson, M. W. (1992). "The Gedankenexperiment Method of Ethics." *The Journal of Value Inquiry*, Vol. 26 (1992): 525-535.
- Janis, Allen I. (1991). "Can Thought Experiments Fail?" In Horowitz and Massey (1991): 113-118.
- Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination and Reason*. Chicago: Chicago University Press.
- Johnston, Mark (1987). "Human Beings." *Journal of Philosophy*, Vol. 84, No. 2 (February 1987): 59-83.
- Johnston, Mark (1989a). "Fission and the Facts." In Tomberlin (1989): 369-397.
- Johnston, Mark (1989b). "Relativism and the Self." In Krausz (1989): 441-472.
- Johnston, Mark (1992). "Reasons and Reductionism." *Philosophical Review*, Vol. 101, No. 3 (July 1992): 589-618.

Johnston, Mark (1997). "Human Concerns Without Superlative Selves." In Dancy (1997): 149-179.

King, Peter (1991). "Mediaeval Thought Experiments: The Metamethodology of Mediaeval Science." In Horowitz and Massey (1991): 43-64.

Kolak, Daniel (1993). "The Metaphysics and Metapsychology of Personal Identity: Why Thought Experiments Matter in Deciding Who We Are." *American Philosophical Quarterly*, Vol. 30, No. 1 (January 1993): 39-53.

Kornblith, Hilary (1996). "The Role of Intuition in Philosophical Inquiry: An Account with No Unnatural Ingredients." In DePaul and Ramsey (1998): 129-142.

Koyré, Alexandre (1968). *Metaphysics and Measurement*. London: Chapman and Hall.

Krausz, Michael, ed. (1989). *Relativism: Interpretation and Confrontation*. Notre Dame: University of Notre Dame Press.

Krimsky, Sheldon (1970). "The Nature and Function of 'Gedankenexperimente' in Physics." Ph.D. dissertation, University of Michigan.

Krimsky, Sheldon (1973). "The Use and Misuse of Critical Gedankenexperimente." *Zeitschrift für allgemeine Wissenschaftstheorie*, Vol. 4 (1973): 323-334.

Kuhn, Thomas (1964). "A Function for Thought Experiments." Reprinted with new pagination in Kuhn (1977).

Kuhn, Thomas (1970). *The Structure of Scientific Revolutions* (second edition). Chicago: University of Chicago Press.

Kuhn, Thomas (1977). *The Essential Tension*. Chicago: University of Chicago Press.

Kujundzic, Nebojsa (1992). "How Does the Laboratory of the Mind Work?" *Dialogue*, Vol. 32, No. 3 (Summer 1992): 573-578.

Kujundzic, Nebojsa (1995). "Thought Experiments: Architecture and Economy of Thought." *The Journal of the British Society for Phenomenology*, Vol. 26, No. 1 (January 1995): 86-93.

Laymon, Ronald (1991). "Thought Experiments of Stevin, Mach and Gouy: Thought Experiments as Ideal Limits and as Semantic Domains." In Horowitz and Massey (1991): 167-192.

Lennox, James G. (1991). "Darwinian Thought Experiments: A Function for Just-So Stories." In Horowitz and Massey (1991): 223-245.

Leplin, Jarrett (1987). "The Role of Experiment in Theory Construction." *International Studies in the Philosophy of Science*, Vol. 2 (Autumn 1987): 72-83.

Lipton, P. "Review of Horowitz and Massey, eds. Thought Experiments in Science and Philosophy (1991)." *Ratio* (new series) 6: 82-86.

Mach, Ernst (1926/1976). *Knowledge and Error* (fifth edition). Trans. Thomas McCormack and Paul Foulkes.

Mach, Ernst (1933/1960). *The Science of Mechanics* (ninth edition). Trans. Thomas McCormack.

MacLachlan, J. (1973). "A Test of an 'Imaginary' Experiment of Galileo's." *Isis*, Vol. 64 (1973): 374-379.

Madell, Geoffrey (1985). "Derek Parfit and Greta Garbo." *Analysis*, Vol. 45 (March 1985): 105-109.

Maffie, James (1997). "Just-So Stories about 'Inner Cognitive Africa': Some Doubts about Sorensen's Evolutionary Epistemology of Thought Experiments." *Biology and Philosophy*, Vol. 12, No. 2 (April 1997): 207-224.

Maidens, Anna (1998). "An Epistemological Role for Thought Experiments." In Shanks (1998).

Martin, Raymond (1993). "Real Values: Why the Wilkes-Donagan Prohibition is Mistaken." *Metaphilosophy*, Vol. 24, No. 4 (October 1993): 400-406.

Martin, Raymond (1995). "Fission Rejuvenation." *Philosophical Studies*, Vol. 80, No. 1 (October 1995): 17-40.

Martin, Raymond, John Barresi, Alessandro Giovannelli, eds. (1998). "Fission Examples in the Eighteenth and Early Nineteenth Century Personal Identity Debate." *History of Philosophy Quarterly*, Vol. 15, No. 3 (July 1998): 323-348.

Massey, Barbara (1991). "Do All Rational Folk Reason as We Do? Frege's Thought Experiment Reconsidered." In Horowitz and Massey (1991): 99-110.

Massey, Gerald (1991). "Backdoor Analyticity." In Horowitz and Massey (1991): 285-296.

Massey, Gerald (1995). "Gedankenexperiment: Where Science and Philosophy Meet." Unpublished manuscript.

McAllister, James W. (1996). "The Evidential Significance of Thought Experiment in Science." *Studies in History and Philosophy of Science*, Vol. 27, No. 2 (June 1996): 233-250.

McCullough, G. (1995). *The Mind and Its World*. London: Routledge Press.

McDowell, John (1994). *Mind and World*. Cambridge: Harvard University Press.

McMullin, Ernan (1978). "The Conception of Science in Galileo's Work." In Butts and Pitt (1978): 209-257.

McMullin, Ernan (1985). "Galilean Idealization." *Studies in History and Philosophy of Science*, Vol. 16 (1985): 247-273.

Miller, Fred D. and Nicholas Smith, eds. (1989). *Thought Probes: Philosophy Through Science Fiction Literature*. Englewood Cliffs, NJ: Prentice Hall.

Mills, Eugene (1993). "Dividing without Reducing: Bodily Fission and Personal Identity." *Mind*, Vol. 102, No. 405 (January 1993): 37-51.

Miscevic, Nenad (1992). "Mental Models and Thought Experiments." *International Studies in the Philosophy of Science*, Vol. 6, No. 3 (1992): 215-226.

Miscevic, Nenad (1997). "Categorial and Essentialist Intuitions: A Naturalist Perspective." *Acta Analytica* (1997): 21-39.

Mohanty, J. N. "The Method of Imaginative Variation in Phenomenology." In Horowitz and Massey (1991): 261-272.

Moor, J. (1982). "Split Brains and Atomic Persons." *Philosophy of Science*, Vol. 49 (1982): 91-106.

Myers, C. Mason (1968). "Thought Experiments and Secret Stores of Information." *International Philosophical Quarterly*, Vol. 8 (June 1968): 180-192.

Myers, C. Mason (1986). "Analytical Thought Experiments." *Metaphilosophy*, Vol. 17, Nos. 2 and 3 (April/June 1986): 109-118.

Naylor, R. H. (1989). "Galileo's Experimental Discourse" in Gooding, Pinch and Schaffer (1989): 117-134.

Nersessian, Nancy (1984). *Faraday to Einstein: Constructing Meaning in Scientific Theories*. Dordrecht and Boston: Kluwer.

Nersessian, Nancy (1992). "How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science." In Giere (1992): 3-44.

Nersessian, Nancy (1993). "In the Theoretician's Laboratory: Thought Experiment as Mental Modeling." *Proceedings of the Philosophy of Science Association*, Vol. 2 (1993): 291-301.

Norton, John (1991). "Thought Experiments in Einstein's Work." In Horowitz and Massey (1991): 129-148.

Norton, John (1996). "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy*, Vol. 26, No. 3 (September 1996): 333-366.

Nozick, Robert (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

Nozick, Robert (1981). *Philosophical Explanations*. Cambridge, MA: Belknap Press.

Oaklander, L. Nathan (1988). "Shoemaker on the Duplication Argument: Survival and What Matters." *Australasian Journal of Philosophy*, Vol. 66 (June 1988): 234-239.

Oderberg, David S. (1989). "Johnston on Human Beings." *Journal of Philosophy*, Vol. 86, No. 3 (March 1989): 137-141.

Olson, David R. and Nancy Torrance, eds. (1993). *Modes of Thought: Explorations in Culture and Cognition*. Cambridge: Cambridge University Press.

O'Neill, Onora (1986). "The Power of Example." *Philosophy*, Vol. 61 (1986): 5-29.

Palmer, Anthony (1984). "The Limits of AI: Thought Experiments and Conceptual Investigations." In Torrance (1984): 43-50.

Parfit, Derek (1971). "Personal Identity." *Philosophical Review*, Vol. 80, No. 1 (January, 1971). Reprinted with new pagination in Perry (1975a): 199-223.

Parfit, Derek (1982). "Personal Identity and Rationality." *Synthese*, Vol. 53 (November 1982): 227-24.

Parfit, Derek (1984/1987). *Reasons and Persons*. Oxford: Clarendon Press.

Parfit, Derek (1995). "The Unimportance of Identity." In Harris (1995b): 13-47.

Passmore, John (1965). "Everything Has Just Doubled in Size." *Mind*, Vol. 74 (1965): 257.

Perry, John (1972). "Can the Self Divide?" *Journal of Philosophy*, Vol. 69, No. 16 (September 7, 1972): 463-488.

Persson, Ingmar (1992). "The Indeterminacy and Insignificance of Personal Identity." *Inquiry*, Vol. 35, No. 2 (June 1992): 249-269.

Piaget, Jean (1946/1970). *The Child's Conception of Movement and Speed*. Trans. G.E.T. Holloway and M.J. Mackenzie. NY: Basic Books.

Plantiga, Alvin (1974). *The Nature of Necessity*. Oxford: Oxford University Press.

Pogue, J. E. (1993). "Identity, Survival, and the Reasonableness of Replication." *Southern Journal of Philosophy*, Vol. 31, No. 1 (Spring 1993): 45-70.

- Poincaré, Henri (1905/1952). *Science and Hypothesis*. New York: Dover.
- Popper, Karl (1959/1992). *The Logic of Scientific Discovery*. Trans. Karl Popper. New York: Routledge.
- Prudovsky, Gad (1989). "The Confirmation of the Superposition Principle: The Role of a Constructive Thought Experiment in Galileo's *Discorsi*." *Studies in History and Philosophy of Science*, Vol. 20, No. 4 (December 1989): 453-468.
- Putnam, Hilary (1975). "The Meaning of Meaning." In Putnam (1975b): 215-271.
- Putnam, Hilary (1983a). "Analyticity and Apriority: Beyond Wittgenstein and Quine." In Putnam (1983b): 115-138.
- Putnam, Hilary (1983b). *Realism and Reason: Philosophical Papers, Volume 3*. Cambridge: Cambridge University Press.
- Putnam, Hilary (1983c). "There is at Least One A Priori Truth." In Putnam (1983b): 98-114.
- Putnam, Hilary (1990). "Is Water Necessarily H₂O?" In Putnam (1990b): 54-79.
- Putnam, Hilary (1994). "Rethinking Mathematical Necessity." In Putnam (1994b): 245-263.
- Putnam, Hilary (1994b). *Words and Life*. Cambridge: Harvard University Press.
- Quine, W.V. (1972). "Review of Identity and Individuation." *Journal of Philosophy*, Vol. 69, No. 16 (September 7, 1972): 488-497.

- Ray, Carolyn (1998). *Identity and Universals: A Conceptualist Approach to Logical, Metaphysical, and Epistemological Problems of Contemporary Identity Theory*. Unpublished doctoral dissertation. Indiana University at Bloomington.
- Rehder, Wulf (1980). "Thought-Experiments and Modal Logics." *Logique et Analyse*, Vol. 23 (December 1980): 407-417.
- Rescher, Nicholas (1991). "Thought Experiment in Pre-Socratic Philosophy." In Horowitz and Massey (1991): 31-41.
- Ritter, Joachim (1974). "Gedankenexperiment." *Historisches Wörterbuch der Philosophie*. Stuttgart: Schwabe & Co.
- Roese, Neal J. and James M. Olson (1995). *What Might Have Been: The Social Psychology of Counterfactual Thinking*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Rotert, Michelle Marie (1992). *Thought Experiments and Philosophical Analysis: An Examination of the Method of Counterexample and the Legitimacy of Appeals to Pre-Theoretical Intuitions*. Unpublished doctoral dissertation, University of Iowa.
- Rovane, Carol (1990). "Branching Self-Consciousness." *Philosophical Review*, Vol. 99, No. 3 (July 1990): 355-395.
- Rovane, Carol (1999). *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton, NJ: Princeton University Press.
- Schechtman, Marya (1990). "Personhood and Personal Identity." *Journal of Philosophy*, Vol. 77, No. 2 (February 1990): 71-92.

- Schechtman, Marya (1996). *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
- Schick, Theodore, Jr. and Lewis Vaughn (1999). *Doing Philosophy: An Introduction Through Thought Experiments*. Mountain View, CA: Mayfield Publishing Company.
- Segre, M. (1980). "The Role of Experiment in Galileo's Physics." *Archive for History of Exact Sciences*, Vol. 23 (1980): 227-252.
- Shafir, Eldar (1998). "Philosophical Intuitions and Cognitive Mechanisms." In DePaul and Ramsey (1998): 59-74.
- Shanks, Niall, ed. (1998). *Idealization IX: Idealization in Contemporary Physics*. Amsterdam: Rodopi.
- Shapin, S. and Schaffer, S. (1985). *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton, NJ: Princeton University Press.
- Shea, W. R. (1972). *Galileo's Intellectual Revolution*. London: Macmillan.
- Shepard, Roger (1994). "Mind and World: From Newton, Einstein and Darwin to Principles of Mind." Unpublished William James Lectures, Harvard University, Fall 1994.
- Shepard, Roger and Lynn Cooper, eds. (1982). *Mental Images and Their Transformations*. Cambridge: Cambridge University Press
- Shoemaker, Sydney (1963). *Self-Knowledge and Self-Identity*. Ithaca: Cornell University Press.
- Shoemaker, Sydney (1985). "Critical Notice of Reasons and Persons." *Mind*, Vol. 94 (1985): 443-453

- Shoemaker, Sydney (1994). "The First-Person Perspective." *Proceedings of the American Philosophical Association*, Vol. 68, No. 2 (November 1994): 7-22.
- Shoemaker, Sydney (1997a). "Parfit on Identity." In Dancy (1997): 135-148.
- Shoemaker, Sydney (1997b). "Self and Substance." In Tomberlin (1997).
- Shoemaker, Sydney and Richard Swinburne (1984). *Personal Identity: Great Debates in Philosophy*. Oxford: Basil Blackwell.
- Slote, Michael (1991). "Review of Sorensen." *Noûs*, (1992). 328-333.
- Smith, Edward E. and Douglas L. Medin (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.
- Snowdon, Paul F. (1991). "Personal Identity and Brain Transplants." In Cockburn (1991): 109-126.
- Sorensen, Roy (1991). "Moral Dilemmas, Thought Experiments, and Conflict Vagueness." *Philosophical Studies*, Vol. 63 (September 1991): 291-308.
- Sorensen, Roy (1992a). *Thought Experiments*. New York and Oxford: Oxford University Press.
- Sorensen, Roy (1992b). "Thought Experiments and the Epistemology of Laws." *Canadian Journal of Philosophy*, Vol. 22, No. 1 (March 1992), 15-44.
- Sorensen, Roy (1995). "Sorensen's Reply to Bunzl and Feldman." *Informal Logic*, Vol. 7, No. 3 (Fall 1995): 399-405.
- Sorensen, Roy (1997). "Roy Sorensen's Review of Soren Haggqvist's Thought Experiments in Philosophy." <http://astro.ocis.temple.edu/~souder/thought/sorensen.txt>

Sosa, Ernest (1996). "Minimal Intuition." In DePaul and Ramsey (1998): 257-270.

Souder, Lawrence (undated). Temple University. "A Home Page for Thought Experiments." <http://astro.ocis.temple.edu/~souder/thought/>

Stevin, Simon (1955). *The Principal Works of Simon Stevin. Vol. 1: General Introduction and Mechanics*. Ed. E.J. Dijksterhuis. Amsterdam: C. W. Swets and Zeitlinger.

Stone, Tony and Martin Davies, eds. (1995). *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.

Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.

Teng, Yu-Jen (1992). *An Inquiry Into Thought Experiment*. Unpublished doctoral dissertation, Southern Illinois University at Carbondale.

Thomas, N. J. T. (undated). "Imagination, Mental Imagery, Consciousness, and Cognition: Scientific, Philosophical and Historical Approaches." <http://web.calstatela.edu/faculty/nthomas/index.htm>

Thomason, Sarah (1991). "Thought Experiments in Linguistics." In Horowitz and Massey (1991): 247-257.

Thomson, Judith Jarvis (1971). "A Defense of Abortion." *Philosophy and Public Affairs*, Vol. 1, No. 1 (Fall 1971): 47-66.

Thomson, Judith Jarvis (1986). *Rights, Restitution and Risk*. Cambridge: Harvard University Press.

Thomson, Judith Jarvis (1997). "People and Their Bodies." In Dancy (1997): 202-229.

Tidman, Paul (1994). "Conceivability as a Test for Possibility." *American Philosophical Quarterly*, Vol. 31, No. 4 (October 1994): 297-309.

Tiles, James E. (1993). "Experiment as Intervention." *British Journal for the Philosophy of Science*, Vol. 44, No. 3 (September 1993): 463-475.

Torrance, Steven, ed. (1984). *The Mind and the Machine: Philosophical Aspects of Artificial Intelligence*. Chichester: Horwood.

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

Urbaniec, Jacek (1988). "In Search of a Philosophical Experiment." *Metaphilosophy*, Vol. 19 (July-August 1988): 294-306.

Vaihinger, Hans (1924). *The Philosophy of "As if"*. Trans. C.K. Ogden. London: Routledge & Kegan Paul.

Ward, David E. (1995). "Imaginary Scenarios, Black Boxes and Philosophical Method." *Erkenntnis*, Vol. 43, No. 2 (September 1995): 181-198.

Wilkes, Kathleen V. (1988). *Real People: Personal Identity without Thought Experiments*. Oxford: Clarendon Press.

Williams, Bernard (1970/1973). "The Self and the Future." Reprinted (with new pagination) in Williams (1973): 46-63. Also in Perry (1975a): 179-198.

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

Wiser, Marianne and Susan Carey (1983). "When Heat and Temperature Were One." In Gentner and Stevens (1983): 267-297.

Witt-Hansen, Johannes (1976). "H.C. Orsted, Immanuel Kant, and the Thought Experiment." *Danish Yearbook of Philosophy*, Vol. 13 (1976): 48-65.

Wittgenstein, Ludwig (1953). *Philosophical Investigations*. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1961). *Tractatus Logico-Philosophicus*, translation by Pears and McGuinness, Routledge & Kegan Paul, London.

Wittgenstein, Ludwig (1967). *Zettel*. Basil Blackwell, Oxford.

Yablo, Stephen (1993). "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research*, Vol. 53, No.1 (March 1993): 1-42.

Index

(prepared by Dave Bzdak)

- accidental regularities, 12
- Achinstein, Peter, 35n
- Ackrill, J.L., 13n, 40n, 103n
- actual cases (and their relation to imaginary cases), 15-18
- All Soul's College (example of), 144n
- Aristotelian response to Galilean thought experiment, 42-45, 150-52
- Aristotelian theory of falling bodies, 33, 39, 42
- Aristotle, 13n, 39, 40n, 48n, 103n
 - On Falling Bodies*, 39
 - On the Heavens*, 40n
 - Physics*, 13n, 40n
 - Metaphysics*, 103n
- Armstrong, David, 13n
- artifacts (as interest-dependent), 77, 84
- Aune, Bruce, 113n
- Bach cello replica (example), 139-42
- Baillie, James, 23n, 113n
- Bealer, George, 20, 34n
- Beck, Simon, 113n
- Brody, Baruch, 157n
- Brown, James Robert, 13n, 18n, 34n, 37n, 38n, 56-62, 151
- Burke, Michael B., 69n
- Butts, Robert E., 39n
- Cambridge changes, 77-78, 77n, 78n
- Carey, Susan, 27n
- Cargile, James, 38n
- Carrier, Martin, 34n
- Cavell, Stanley, 53n
- Chisholm, Roderick, 87n
- Claggett, Marshall, 39n
- Clavelin, Maurice, 39n
- Clement, John, 38n
- closest continuer theory of identity
 - See also* Nozick, Robert
 - characterization of, 92-97
 - counterintuitive implications of, 93-97
- Cockburn, David, 113n
- Cooper, Lane, 39n

- counterfactual cases 17-18, 17n, 35, 156-57
See also hypothetical cases
- Damerow, Peter, 39n
- Dancy, Jonathan, 14-15, 15n
- dead body case (example), 138-39
- DePaul, Michael R., 18n
- Derivativity Thesis (and factive thought experiments)
 characterization of, 34, 37-39
 criticism of, 49-56, 63
- Descartes, René, 145
- Diamond, Cora, 144n
- Dijksterhuis, E.J., 39n
- Dispensability Thesis (and factive thought experiments)
 characterization of, 34, 37-39, 42
 criticism of, 47-50, 63
 and Galileo's thought experiment, 42
- Donagan, Alan, 113n
- Drabkin, I.E., 39n
- Drake, Stillman, 39n, 40n, 43n
- Einstein, Albert, 19, 24n, 25, 29-30, 151n
 and clock-in-a-box thought experiment, 24n
 and debates with Neils Bohr, 24n
 and moving train thought experiments, 19, 29-30
- Einstein-Podolsky-Rosen thought experiment, 151n
- Elimination Thesis (and factive thought experiments)
 characterization of, 34-36

- negative argument for, 36-37
 rejection of, 53-56
- positive argument for, 36-37
 rejection of, 49-53
- entification, 45, 52
- entity-creating processes, 69-70, 81, 88, 107, 117, 146, 155
- epistemology (and imaginary cases), 14-15
- Ethics*, 113n
- exceptional cases
 characterization of, 2-8
 and conceptual commitments, 12
 examples of, 2-5, 6-8
 and imaginary scenarios, 1, 12
 and the Ship of Theseus thought experiment, 70-79, 98, 109
- strategies for accounting for, 8-12
 exception-driven norm strategy, 8-12
 norm-driven exception strategy, 8-12
 as test cases, 12
- in theories with privileged characteristics, 2-5
- in theories without privileged characteristics, 6-8
- and thought experiments, 12-14, 142, 152-59
- exception-driven norms, 8-12
- experience machine thought experiment, 19, 21, 24-25, 30, 132
- experiments, scientific, 27

- Fodor, Jerry A., 7n
- four-dimensionalism, 70n
- Gale, Richard, 113n
- Galileo, 33, 38-42, 46-48, 52, 54n, 57, 61, 78n, 156
Discourse Concerning Two New Sciences, 39
- Galileo's thought experiment about falling bodies, 26, 33, 38-63
- argumentative reconstruction of, 41-42
 criticism of, 45-47
- Aristotelian response to, 42-48, 150-52
- Brown's discussion of (platonist account), 56-62, 151
- constructivist account of, 56-62, 151
- description of, 39-41
 how it leads to new beliefs/knowledge, 54-55, 151-54
- Norton's discussion of (empiricist account), 36-37, 56-62, 150
- Garrett, Brian, 69n, 119n
- Geach, Peter, 77n
- Gedankenexperiment* 20n, 55
See also thought experiment
- Gelman, Susan, 27n
- Gendler, Tamar Szabó, 113n
- Gentner, Dedre, 27n
- Goldstone, Robert L., 7n
- Gooding, David C., 34n
- Goodman, Nelson, 140n, 141n
- Hacking, Ian, 34n
- Häggqvist, Soren, 18n
- Haksar, Vinit, 113n
- Hannaway, Owen, 35n
- "hard cases make bad law" adage, 12
- "hard cases make good science" adage, 12-13
- Hawthorne, John, 114n
- Hertzberg, Lars, 23n, 113n
- Hesse, Mary, 35n
- Hirsch, Eli, 68n, 77n, 88-91, 101n, 103n, 108n, 154n, 155n
- on identity and the Ship of Theseus thought experiment, 88-91
- Compositional Criterion, 89-91, 108n
- response to, 90-91
- Sortal Rule, 89-91
- Hirschfeld, Lawrence, 27n
- Hobbes, Thomas, 67
- Horowitz, Tamara, 18n, 35n
- Humean uniformities (or regularities), 13n
- Humphreys, Paul, 34n
- hypothetical cases
 characterization of, 14-18, 17n, 20, 35, 154
See also counterfactual cases
- idealization (in imaginary cases), 17
- identity
 of artifacts, 71-73, 76-85
 automatic vs. specially-secured, 70-72
- candidate sufficient principles for, 99-109

- identity *continued*
 criteria of, 65-72, 98, 101-03
 and disruptions in spatio-temporal continuity, 68-69
 extrinsically determined, 66-67, 103, 156
 intrinsically determined, 66-67
 Locke's general theory of, 71
 logic of, 70-71, 75
 of organisms, 71-80
 under a sortal, 82-84
 identity-preserving processes, 69-70, 72n, 81, 88, 90, 103, 107, 109, 117, 146
 imaginary cases (or scenarios)
 and actual cases, 15-18
 and actual experiments, 112n
 and epistemology, 14-15, 25
 and exceptional cases, 1, 12
 and factive thought experiments, 33
 puzzle of, 1
 utility of, 12-17, 52-55
 intuitions, 20-21
 physical, 20-21
 rational, 20-21
 Irvine, Andrew, 34n
 Irwin, Terence, 133n, 139n
 Janis, Allen I., 24n, 34n
 Johnston, Mark, 96n, 113n, 119n, 130-131
 Kant, Immanuel, 11, 11n, 14n
Groundwork for the Metaphysics of Morals, 11
 Keil, Frank, 27n

- kinds of thought experiment
 factive, conceptual and valuational, 25-27, 149-59
passim
 hypothetical and counterfactual, 17-18, 156-57
 Komatsu, Lloyd K., 7n
 Kolak, Daniel, 113n
 Korsgaard, Christine, 144n
 Koyré, Alexandre, 35n, 38n, 39n, 43n
 Kuhn, Thomas, 2-5, 35n, 38n, 48n, 55n, 150n
 Kujundzic, Nebojsa, 34n
 Langton, Rae, 123n
 Laurence, Stephen, 7n, 27n
 laws of nature (in Armstrong), 13n
 Laymon, Ronald, 34n
 Lewis, David, 123n, 124n
 Lipton, P., 34n
 Lloyd, B.B., 7n
 Locke, John, 71, 82n, 144n
 Locke's theory of identity over time, 71
 Lowe, E.J., 69n
 Mach, Ernst, 20n, 35n, 55, 151n
 Mackie, J.L., 70-72, 75-76, 82n, 112n
 Madell, Geoffrey, 23n, 121n
 Margolis, J., 7n, 27n
 Markman, Ellen M., 27n
 Martin, Raymond, 23n, 113n, 115n, 145n
 Massey, Gerald, 18n, 34n, 35n
 McDowell, John, 144n
 Medin, Douglas L., 7n

- Merricks, Trenton, 67n, 69n
 Mervis, C.G., 7n
 Mill, John Stuart, 12, 136-140
 Mill's Methods, 26, 115, 136-137, 138-140, 147
 Mills, Eugene, 124n
 Molnar, G., 13n
 Nagel, Thomas, 119n
 naïve regularity theories, 13, 13n
 Neisser, Ulrich, 7n, 27n
 Nersessian, Nancy, 34n, 54n
New York Times Sunday Magazine, 143n
 Newton, Isaac, 20
 rotating bucket thought experiment, 20
 Noonan, Harold, 69n, 123n, 124n
 normative legal theories, 14
 norm-driven exceptions, 9-12, 153, 156
 Norton, John, 34, 35n, 36-37, 37n, 38n, 41n, 53-54, 56-62, 150-151
 notional entities, 86-87
 Nozick, Robert, 15n, 19, 21, 30, 69n, 73n, 91-97, 97n, 112n, 123n, 132, 139n, 154n, 155n
 on identity and the Ship of Theseus thought experiment, 91-97
 on personal identity, 123n
Philosophical Explanations, 91
 Oderberg, David S., 23n, 113n
 Olson, Eric, 27n, 121n
 O'Neill, Onora, 18n
 Örsted, Hans Christian, 20n
 paradigm cases
 characterization of, 6-8, 6n
 and exceptional cases, 8
 instances of, 7-10
 Parfit, Derek, 66n, 77n, 78n, 85-88, 97, 113-147, 154n, 157-158
 on identity and the Ship of Theseus thought experiment, 85-88
 on personal identity, 113-47
 assumptions of his argument
 intrinsicness principle, 114-15, 123, 129-34, 137, 138-45
 necessity principle, 114-15, 134, 147
 sufficiency principle, 114
 parity principle, 122-23, 129
 context for his argument, 115-20
 fission argument, 120-25, 146-147
 criticisms of, 129-37
 evaluation of, 125-30
 why it is compelling, 134-37
 unimportance of identity conclusion, 125, 128
Reasons and Persons, 113-47
 Penelhum, Terence, 73n
 Perry, John, 112n
 personal identity, 111-47
 and brain-transplant examples, 111, 116

personal identity *continued*
 and cases of qualitative identity
 (or indistinguishability),
 116
 casewise explanatory difference
 principle of, 134-35, 137,
 147
 and exceptional cases, 142-45
 facts of life and, 111-15, 146
 and fission cases, 111, 116-45,
 157-59
 evolutionary issues concerning,
 119-20
 examples of, 118
 problems with, 118-19
 what they might show, 116-
 20
 and fusion cases, 74n
 imaginary cases and, 115-16
 and motherhood (example),
 142-43
 necessary conditions for, 113
 and teletransportation cases,
 111, 116, 117n
 and valuation, 139-45
 philosophical theories (and excep-
 tional cases), 14
 Piaget, Jean, 2-4
 pickle-barrel (example), 104-06
 Pitt, Joseph C., 39n
 Plato, 19
 pledge of allegiance (example), 15
 Popper, Karl, 35n
 privileged characteristics (of theo-
 retical entities or situa-
 tions), 2, 5, 6, 8
 privileged description (of entities
 or situations under a the-
 ory), 2, 5, 6

proxy characteristics (of theoreti-
 cal entities or situations),
 2, 5
 Prudovsky, Gad, 35n, 38n
 Putnam, Hilary, 19, 21, 30, 119n
 Quine, W.V.O, 113n
 Ramsey, William, 18n
 rational prudential concern 113-
 14, 122-39, 143-47
See also personal identity
 Rea, Michael C., 67n, 69n
 ring of Gyges, 19, 25, 28
 Robertson, Teresa, 122n
 Robinson, John, 119n, 124n
 Rosch, Eleanor, 7n
 Rosenberg, Jay F., 77n, 82
 Rovane, Carol, 23n, 113n, 126n,
 133n, 144n
 Rubiner, Michael, 143n
 Saunders, Simon, 153n
 Scaltsas, Theodore, 69n
 Schechtman, Marya, 144n, 145n
 scientific theories, 13-14
 Sedley, David, 76n
 Segre, M., 39n
 Shanks, Niall, 17n, 35n
 Shepard, Roger, 35n, 56n, 151n
 Ship of Theseus thought experi-
 ment, 65-109, 154-56
 description of, 67-68
 puzzles raised by, 68-97
 diagnosis of, 97-109
 solutions to, 79-97
 Shoemaker, Sydney, 76n, 112n,
 121n, 145n
 Sider, Theodore, 66n, 70n, 85n

Smart, Brian, 69n, 108n
 Smith, Edward E., 7n
 Snowdon, Paul F., 113n
 Socrates, 28
 Sorenson, Roy, 18n, 35n, 38n, 56n
 Sosa, Ernest, 113n, 131, 131n
 Stevens, Albert L., 27n
 Stevinus, 19, 20n, 21, 24-26,
 28-29
 Stevinus' inclined plane thought
 experiment, 19, 21, 25,
 28-29
 Strawson, P.F., 21n, 22n
 Sturgeon, Nicholas, 134n
 Sunstein, Cass, 53n
 Swinburne, Richard, 112n
 Szabó, Zoltán Gendler, 84n
 theory of gender (example), 7-11,
 11n
 theory of sex (example), 5-11
 Thomas, N.J.T., 54n
 Thomason, Sarah, 54n
 Thomson, Judith Jarvis, 19, 25,
 31, 54n, 121n
 Thomson's "ailing violinist"
 thought experiment, 19,
 25, 31
 thought experiments
 as actual experiments in
 thought, 54-56, 65-66,
 152, 158
 as disguised arguments, 33-35
 rejection of, 47-53
 characterization of, 18-24, 34
 clarification of terminology
 about, 34-35
 and conceptual frameworks, 51-
 56, 62, 152-59
 constructive participation in,
 54-56
 controversy about terminology
 about, 20-21
 criticisms of
 inapplicability, 22-24, 155
 unimaginability, 22-24, 154
 unsound argument, 22-24,
 154, 157
 demonstrative force of, 33-36,
 50-52
 and imaginary scenarios, 34
 justificatory force of, 56
 kinds of, 25-27
 conceptual, 25-26, 65, 154-56
 factive, 25-26, 33, 150-54
 valuational, 25-26, 156-59
 and new beliefs/knowledge, 47-
 49, 53-56, 151-54
 platonic, 56-62
 and scientific experiments, 27
 six famous examples of, 19, 25,
 28-31
 tripartite structure of, 21-25
 Torrance, Steven, 27n
 "Tortoise and the Hare" (exam-
 ple), 4n
 twin earth, 19, 21, 25, 30
 Unger, Peter, 67n, 103n, 112n,
 113n
 Van Inwagen, Peter, 67n, 73n,
 77n, 79-81, 79n, 80n,
 84-85, 97, 154n, 155n
 on identity and the Ship of
 Theseus thought experi-
 ment, 79-87, 97
Material Beings, 79

Index

- Wellman, H.M., 27n
White, Stephen, 113n
Whiting, Jennifer, 144n
Wiggins, David, 68n, 74n, 82-83,
 84n, 93, 113n, 117n,
 119n, 123n
Wilkes, Kathleen, 23n, 35n, 113n,
 119n
Williams, Bernard, 117n, 121n
Witt-Hansen, Johannes, 20n
Wittgenstein, Ludwig, 7n
Wolf, Susan, 144n