

**Thought Experiments In Ethics:
A Contextualist Approach to the Grounding Problem**

ANNE HARLAND

M.A., University of British Columbia, 1993

B.A., Simon Fraser University, 1988

A.R.C.T., Royal Conservatory of Toronto, 1977

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Philosophy

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October, 1998

© Anne Harland, 1998

Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy
The University of British Columbia
1866 Wesbrook Mall
Canada V6T 1Z1

December 14th, 1998

Abstract

How can an experiment which occurs only in thought lead to new and accurate conclusions about the world beyond thought? What makes thought experiments relevant to the domains they are designed to explore?

One answer is that successful thought experiments are grounded. Explaining the nature of this grounding relationship, especially as it applies to ethics, is the main task of this dissertation.

A thought experiment is an *experiment* that occurs in *thought*. The "thought" label distinguishes it from an ordinary physical experiment, while the "experiment" label distinguishes it from other types of merely analogical, conjectural, or hypothetical reasoning. Many of the components that are necessary for a successful physical experiment are also necessary for a successful thought experiment. A thought experiment, like a physical experiment, must isolate and vary variables in order to answer a question within a given theoretical context. The result of the experiment has repercussions for its theoretical context.

The grounding relationship holds between the components of the thought experiment and the theoretical context of the thought experiment. In order for the thought experiment to be successful, both the experimental set-up and our responses to it need to be grounded in the thought experiment's theoretical context.

An experimental set-up will be grounded whenever it meets the following conditions. The concepts used must be defined normally, dependent and independent variables must be isolated and relevantly related, and the propositions of the thought experiment (excepting those describing extraneous particulars) must be relevantly related to the given theoretical context and the question under examination.

Grounding responses to thought experiments will then be largely a matter of anticipating and disarming distorting influences. Factors influencing responses include the individual's knowledge of the theoretical context, the state of development of that context, the nature of the presentation of the thought experiment, and subjective filters.

It is sometimes difficult to ascertain whether a thought experiment in ethics is grounded. This is largely due to the nature of the theoretical context of thought experiments in ethics. In order to assess the relationship of thought experiments in ethics to their theoretical context, I advocate employing a contextualist methodology involving the process of wide reflective equilibrium. While contextualists use this approach to arrive at considered judgements relating to specific ethical problems, I show that wide reflective equilibrium can also be used to examine the grounding of thought experiments.

I conclude the dissertation with an examination of the relationship of thought experiments to computer simulations, a study of various common thought experiment distortions, and some tests and methods designed to aid constructing successful thought experiments.

Table of Contents

Abstract	ii
Table of Contents	iii
Acknowledgements	v
1. Introduction	i
2. Thought Experiments: Background	12
2.1 What is a Thought Experiment?	12
2.2 Thought Experiments and Arguments	22
2.3 Platonic Perception	28
2.4 Extraneous Particulars and Simplifying Assumptions	37
2.4.1 Extraneous Particulars	37
2.4.2 Simplifying Assumptions	38
2.5 Logical Structure of Thought Experiments	43
2.5.1 Deductive Thought Experiments	43
2.5.2 Arguments by Analogy	48
2.5.3 Inductive Thought Experiments	54
2.5.4 Thought Experiments which Test Explanatory Potential	61
2.5.5 Ampliative Reasoning in Computer Simulations	64
2.6 Non-Thought Experiments	69
3. Thought Experiments: The Grounding Problem	77
3.1 Grounding	77
3.2 Grounding Propositions	83
3.3 Explicating Grounding	85
3.3.1 Concepts	86
3.3.2 Theoretical Contexts	93
4. Grounding Thought Experiments in Ethics	103
4.1 Contextualism	106
4.2 The Design Problem Approach	117
4.3 A Contextualist Approach to Thought Experiments in Ethics	119
5. Grounding Thought Experiments in Various Contexts	126
5.1 Thought Experiments in Philosophy of Language	127
5.2 Thought Experiments Regarding Personal Identity	132
5.3 Thought Experiments in Science	138
5.4 Thought Experiments and Computer Simulations	140
5.4.1 Limits of Simulations	152
5.4.2 Simulations and Thought Experiments Compared	154
6. Grounding Responses	158
6.1 Responses and the Theoretical Context	163
6.2 Subjective Distortions: Personal Identity Thought Experiments	170
6.3 Subjective Distortions: Thought Experiments in Ethics	176

7. Eliciting Grounded Responses.....	183
7.1 Stories and Details	183
7.2 Contextualism, Literature and Detail	189
7.3 Comparative Method	191
8. Contextualism and Thought Experiments.....	201
Bibliography	206
Index of Thought Experiments & Related Examples	216

Acknowledgements

I would like to thank my supervisor, Andrew Irvine, for his insightful guidance of this project and for all that I have learned from him. Andrew has provided me with a steady source of advice and encouragement that has helped to keep me focussed and on track. I'd also like to thank the other members of my committee, Peter Danielson and Jim Dybikowski, for their support and very helpful comments and advice.

Friends have been very encouraging of me and my work. I would like to thank, in particular, Gail Andestad, Coat Cooke, Paul Stein and Makiko Suzuki. Jill McIntosh and Tom Handley each deserve a special thanks. Jill has been a strong source of encouragement and lively philosophical discussion every step of the way. Tom has unfailingly lent his technical, and most importantly, moral support. His calm and optimistic attitude has helped me a great deal.

My family has been very encouraging. I would like to thank my mother Noëlene Harland, my father William Harland, and my siblings — Richard, Bruce, and especially Lorraine.

I would like to thank the government of Canada for a SSHRC Graduate Fellowship from 1993 to 1997 and the University of British Columbia for a University Graduate Fellowship from 1992 to 1993. I also would like to thank the Philosophy Department for giving me Teaching Assistantships from 1992 to 1997.

I feel very fortunate to have had the encouragement of many people in writing this dissertation. It is difficult to express the depth of my appreciation for this support. If I have left anyone out, I hope they know how important their support has been.

1. Introduction

How can an experiment which occurs only in thought lead to new and accurate conclusions about the world beyond thought? What makes thought experiments relevant to the domain they are designed to explore?

One answer is that successful thought experiments are grounded. The grounding relationship holds between the components of the thought experiment and the theoretical context of the thought experiment. The thought experiment is composed of two parts: the experimental set-up and the responses to the experimental set-up. In order for the thought experiment to be successful, each of these parts needs to be grounded in the theoretical context of the thought experiment.

I apply Reichenbach's distinction between the context of discovery and the context of justification to my examination of thought experiments.¹ In the context of discovery, thought experimenters are free to do as they please. In the context of justification, the thought experiment is seen to be successful, or not. Grounding is a necessary condition for a successful thought experiment. Grounding concerns may, or may not, be prominent in the context of discovery, but they are critical to the assessment of the thought experiment in the context of justification.

My account of grounding of the experimental set-up has two parts. First, the propositions comprising the thought experiment must use concepts that are defined normally; and second, the propositions comprising the thought experiment (excepting those that describe extraneous particulars) must be relevantly related to the theoretical context. Here is a brief outline of each point:²

¹ Reichenbach [1938], 6f.

² In chapter 3, this account is presented and defended in depth.

1. A grounded thought experiment uses concepts that are defined normally. A concept that is defined normally is one which can be described by propositions which conform to the theoretical context of the thought experiment. This means that the propositions that describe the concept must be consistent, plausible, and testable with respect to the theoretical context. The theoretical context consists of the primary theory and background theories from which the data from empirical observation or beliefs about values are expressed. An exception to this condition occurs when it is an aspect of the concept itself that is under scrutiny in the thought experiment.
2. In 1., above, the concern is that propositions that describe the concepts used in the thought experiment are grounded in the theoretical context. These propositions are not directly expressed in the thought experiment. Now, we consider the explicit propositions of the thought experiment. A grounded thought experiment uses propositions (excepting those describing extraneous particulars) which are relevantly related to the theoretical context. The conditions that need to obtain for a particular proposition of a thought experiment to be relevantly related to its theoretical context depend on the function of that proposition. There are four functions of propositions employed in thought experiments. Some propositions describe the dependent variable, some the independent variable, some simplifying assumptions and some extraneous particulars. Consider what it means for each of the first three types of propositions to be relevantly related to their theoretical context:

A. The dependent variable: Propositions that express the dependent variable must conform to the theoretical context of the thought experiment. This means that the propositions must be consistent with, relevant to, and plausible with respect to the theoretical context.

B. The independent variable: Although the propositions that express the independent variable need not conform to the propositions that make up the

theoretical context, the independent variable must be varied in a way that is relevantly related to the theoretical context and the question under examination.

C. Simplifying assumptions and idealized assumptions: Simplifying assumptions and idealized assumptions need not conform to the theoretical context, but they must be carried forward to the conclusion.

Extraneous particulars are experimental conditions not affecting the variables. These need not be relevantly related to the theoretical context of the thought experiment.

Besides the experimental set-up, responses to thought experiments must be grounded or tempered. Grounding these is largely a matter of anticipating and disarming distorting influences. Factors influencing responses are the individual's knowledge of the thought experiment's theoretical context, the state of development of that context, the nature of the presentation of the thought experiment, and subjective filters.

Ascertaining whether thought experiments concerning ethics are grounded can be difficult. Since grounding concerns the relationship between the thought experiment and its theoretical context, I consider the nature of the theoretical context of thought experiments in ethics and how this nature affects the grounding. I call the primary theory of most thought experiments concerning ethics a "framework of values." A framework of values is a set of articulated and unarticulated values held by an individual. It is a developing and loosely defined framework held by a particular person at a particular time. Each framework of values consists of values and norms which have accrued over the history of human social experience, which are often considered to be values and norms of common sense, and values related to one's understanding of oneself, as situated in the natural and the social world. Responses to thought experiments arise from frameworks of values. Through assessing and comparing various responses, the conclusion of the thought experiment is obtained.

This dissertation shows how the contextualist conception of wide equilibrium can be applied to the assessment of thought experiments in ethics. Ethical problems can be seen as design problems whose solutions are formed with the goal of adequately meeting the multiple constraints acting on the situation. The process of wide equilibrium is a process of balancing those constraints. The employment of the wide equilibrium process in an analysis of the thought experiment helps in ascertaining whether the thought experiment is grounded.

My choice to deploy the contextualist methodology in the examination of thought experiments in ethics is a pragmatic one. Thought experiments in ethics are conducted against a variety of contexts and a variety of frameworks of values. For example, some may use thought experiments to explore and develop utilitarian theory, some evolutionary ethics, some an instrumentalist view, and others a contextualist framework. It is hoped that the account given here will be applicable to all thought experiments in ethics. Remember that grounding concerns the relationship of the thought experiment to its theoretical context. The process of examining this relationship is the same, whatever that particular theoretical context may be. In this process the contextualist methodology of wide reflective equilibrium is a useful tool. The moral realist, the intuitionist, the utilitarian and others should find the account given here useful for examining the relationship of their thought experiments to their particular theoretical context.

Although my main concern is the grounding of thought experiments in ethics, throughout the thesis I use thought experiments in various fields to illustrate my account and to serve as comparisons for the more difficult and less thoroughly examined thought experiments in ethics. I begin here with an example of a thought experiment in physics.

Consider the following thought experiment, given by Galileo. Through examining it, we see the manner in which the various components of thought experiments can be understood and, at the same time, become familiar with the terminology used

throughout this dissertation. The original thought experiment is found in Galileo's *Two New Sciences* in which he outlines his physics in dialogue form. The three characters in the dialogue are called *Salviati*, *Sagreto* and *Simplicio*. *Salviati* speaks for Galileo, *Sagreto* is *Salviati*'s foil and prod, and *Simplicio* is named after *Simplicius*, the famous Greek commentator on Aristotle. Although "*Simplicio*" is the Italianate translation of *Simplicius*, it means "simpleton" in Italian. Here is the thought experiment:

Salviati: But, even without further experiment, it is possible to prove clearly, by means of a short and conclusive argument, that a heavier body does not move more rapidly than a lighter one provided both bodies are of the same material and in short such as those mentioned by Aristotle. But tell me, *Simplicio*, whether you admit that each falling body acquires a definite speed fixed by nature, a velocity which cannot be increased or diminished except by the use of force [violenza] or resistance.

Simplicio: There can be no doubt but that one and the same body moving in a single medium has a fixed velocity which is determined by nature and which cannot be increased except by the addition of momentum [impeto] or diminished except by some resistance which retards it.

Salviati: If then we take two bodies whose natural speeds are different, it is clear that on uniting the two, the more rapid one will be partly retarded by the slower, and the slower will be somewhat hastened by the swifter. Do you not agree with me in this opinion?

Simplicio: You are unquestionably right.

Salviati: But if this is true, and if a large stone moves with a speed of, say, eight while a smaller stone moves with a speed of four, then when they are united, the system will move with a speed less than eight; but the two stones when tied together make a stone larger than that which before moved with a speed of eight. Hence the heavier body moves with less speed than the lighter; an effect which is contrary to your supposition. Thus you see how, from your assumption that the heavier body moves more rapidly than the lighter one, I infer that the heavier body moves more slowly.

Simplicio: I am at sea because it appears to me that the smaller stone when added to the larger increases its weight and by adding weight I do not see how it can fail to increase its speed or, at least, not to diminish it.

Salviati: Here again you are in error, *Simplicio*, because it is not true that the smaller stone adds weight to the larger.

Simplicio: This is, indeed, quite beyond my comprehension.

Salviati: It will not be beyond you when I have once shown you the mistake under which you are laboring. Note that it is necessary to distinguish between heavy bodies in motion and the same bodies at rest.

A large stone placed in a balance not only acquires additional weight by having another stone placed upon it, but even by the addition of a handful of hemp its weight is augmented six to ten ounces according to the quantity of hemp. But if you tie the hemp to the stone and allow them to fall freely from some height, do you believe that the hemp will press down upon the stone and thus accelerate its motion or do you think that motion will be retarded by a partial upward pressure? One always feels the pressure upon his shoulders when he prevents the motion of a load resting upon him; but if one descends just as rapidly as the load would fall how can it gravitate or press upon him? Do you not see that this would be the same as trying to strike a man with a lance when he is running away from you with a speed which is equal to, or even greater, than that with which you are following him? You must therefore conclude that, during free and natural fall, the small stone does not press upon the larger and consequentially does not increase its weight as it does when at rest.

Simplicio: But what if we should place the larger stone upon the smaller?

Salviati: Its weight would be increased if the larger stone moved more rapidly; but we have already concluded that when the small stone moves more slowly it retards to some extent the speed of the larger, so that the combination of the two, which is a heavier body than the larger of the two stones, would move less rapidly, a conclusion which is contrary to your hypothesis. We infer therefore that large and small bodies move with the same speed provided they are of the same specific gravity.

Simplicio: Your discussion is really admirable; yet I do not find it easy to believe that a bird-shot falls as swiftly as a cannon ball.³

Many of the components that are necessary for the construction of a successful physical experiment are also necessary for the construction of a successful thought experiment. For example, through Galileo's thought experiment we see that a thought experiment, like a physical experiment, isolates and varies variables in order to answer a well-formed question within a given theoretical context. The result of the experiment has repercussions for its theoretical context. In this case, the dependent variable is the rate of velocity of the falling object(s) and the independent variable is the weight of the falling object(s). The independent variable is varied, and its effect on the dependent variable is tracked. In Galileo's thought experiment the light object, the heavy object, and the light

³ Galileo [1638], 60f.

and heavy objects tied together, are considered in relation to the rate of descent of these objects.

Thought experiments operate in relation to a theoretical context. The theoretical context consists of the primary theory and background theories. Here, the primary theory is the theory of physics: the thought experiment is concerned with the effect of weight on falling objects. One of the background theories is classical logic: the thought experiment generates a contradiction; we see that the light and heavy objects tied together cannot fall both more quickly and more slowly than the heavy object alone. We realize that this contradiction shows that the premiss of the thought experiment, Aristotle's hypothesis about the relationship between the weight and velocity of falling objects, must be incorrect.

Thought experiments also employ simplifying assumptions which are used to help isolate the variables. In this case, an unstated assumption is that the falling objects are not affected by friction. Another feature of thought experiments is the presence of extraneous particulars. In this case, the dialogue form and some of the conversation between the characters is extraneous.

Although I began here with Galileo's thought experiment, which is a scientific thought experiment, recall that my main concern is thought experiments in ethics. It is useful to begin with Galileo's thought experiment for two reasons.⁴ First, the comparisons between it and a physical experiment are readily apparent. For example, through the Galileo experiment we are able to examine the various parts, and the functions of various parts, of thought experiments in relation to the manner in which they would be considered in a physical experiment. Throughout the thesis, I employ the well-developed analysis of physical experiments as a clear account against which to compare

⁴ I discuss this thought experiment further in §2.3, §2.4.1, §2.4.2 and §2.5.1.

and contrast thought experiments. Second, since the grounding relationship is one that holds between the components of a thought experiment and their theoretical context, the extent to which a thought experiment is grounded is easier to assess when the theoretical context is more clearly defined. The theoretical context of science is more well-developed than that of ethics. For this reason, I use the more easily analyzed and assessed scientific thought experiments as a touchstone for comparison with thought experiments in ethics throughout the thesis.

Now, consider a well-known thought experiment in ethics:

... let me ask you to imagine this. You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you, 'Look, we're sorry the Society of Music Lovers did this to you — we would never have permitted it if we had known. But still, they did it, and the violinist now is plugged into you. To unplug him would be to kill him. But never mind, it's only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from you.' Is it morally incumbent on you to accede to this situation? No doubt it would be very nice of you if you did, a great kindness. But do you have to accede to it? What if it were not nine months, but nine years? Or longer still? What if the director of the hospital says, 'Tough luck, I agree, but you've now got to stay in bed, with the violinist plugged into you, for the rest of your life. Because remember this. All persons have a right to life and violinists are persons. Granted you have a right to decide what happens in and to your body, but a person's right to life outweighs your right to decide what happens in and to your body. So you cannot ever be unplugged from him.'⁵

This thought experiment employs an argument by analogy. Thomson argues that the situation of being kidnapped and hooked up to an ill violinist is similar to that of being pregnant due to rape, and so our judgements about the one case should be applied to those about the other. Analogies must be used with care. In particular, one must

⁵ Thomson [1971], 38-39.

consider whether there are confounding variables which make the situations in question disanalogous.⁶

Thought experiments must be amenable to characterization as arguments or sets of arguments. In §2.5.1, I present the ill-violinist thought experiment in argument form. From the argument form we can ascertain what kind of reasoning is employed in the thought experiment. The argument can be evaluated for soundness, in the case of deductive reasoning, or strength, in the case of inductive reasoning.

While some thought experiments argue by analogy, others test the explanatory potential of a theory. Consider Darwin's thought experiments in *On the Origin of Species*. Darwin's thought experiments test the explanatory potential of his theory by describing plausible situations in which variations in species would occur in the way presented by the theory. These thought experiments are designed to test the theory's ability to explain certain phenomena, rather than the truth of the statements of the theory.⁷ The explanations tested by this type of thought experiment need to be assessed. An explanation must be consistent with relevant background theories and plausible with respect to the evidence. Then, the explanation needs to be ranked in relation to other possible explanations. Successful thought experiments must not only be grounded, but also employ strong reasoning or provide good explanations.

My account of grounding applies to thought experiments concerning areas other than ethics. In Chapter 5, I show how it applies to thought experiments in philosophy of language, philosophy of personal identity and science. I close chapter 5 with an examination of computer simulations. Some types of computer simulations are like thought experiments. I call these "simulation experiments". Simulation experiments either test the explanatory potential of a theory, or test the causal links between one state

⁶ Arguments by analogy are considered in §2.5.2.

⁷ Lennox [1991], 229.

of affairs and another. For example, simulations of birds flocking test the explanatory potential of a certain theory about flocking birds which is modeled by that simulation.

Thought scenarios such as Schrödinger's Cat and Maxwell's Demon, although widely referred to as thought experiments, fail to meet my criteria for that label.⁸ These thought scenarios illustrate a theory rather than test a question relative to a theoretical context. In the same way, some simulations are not *simulation experiments*. I call these "*simulation demonstrations*". Flight simulators are used as tools to understand and interact with modeled states of affairs that are constructed from well-established theories. They do not test the explanatory potential of a theory, or test links between a proposed theory and a state of affairs.

I close the dissertation with an outline of a constructive approach for developing grounded thought experiments in ethics. Supplying relevant detail to thought experiments and considering variations of thought experiments helps to ground the thought experiment and temper one's responses. I present, and consider the merits of, a comparative method which employs these strategies for constructing successful thought experiments in Chapter 7.

The approach to the problem of grounding thought experiments in ethics that motivates the account given here is reliabilist. I am seeking a way to construct thought experiments in ethics that will be reliably useful to us and relevant to our beliefs about values. The contextualist approach to ethics supports my account. Contextualists apply the idea of wide reflective equilibrium to the ethical sphere. I show that we can use the wide reflective equilibrium process to examine the strength of the relationship of the thought experiment to its theoretical context.

⁸ See §2.6.

Thought experiments that are successful share many features with successful physical experiments. In addition, they are adequately related to their theoretical context. When a thought experiment is adequately related to its theoretical context it is grounded. In this dissertation, I give an account of what would be required for the two components of thought experiments — the experimental set-up and our responses — to be grounded.

2. Thought Experiments: Background

2.1 What is a Thought Experiment?

One need look no further than the components of the term "thought experiment" for its definition. As Sorensen writes:

A thought experiment is an experiment that purports to achieve its aim without the benefit of execution.⁹

In other words, a thought experiment must both be something that occurs in thought and also something that qualifies for the label "experiment". An experiment, in turn, may be understood to be

... a procedure for answering or raising a question about the relationship between variables by varying one (or more) of them and tracking any response by the other or others.¹⁰

The "thought" label distinguishes a thought experiment from an ordinary physical experiment, while the "experiment" label distinguishes a thought experiment from an example of merely analogical, conjectural, or hypothetical reasoning.

A thought experiment, then, is an *experiment* that occurs in *thought*. In this section, I concentrate on the *experiment* part of thought experiments. In the next chapter, I consider the *thought* part.

In order to understand how experiments work, it helps to consider particular experiments. Consider the following simple experiment: Suppose we want to know under what conditions a certain type of plant thrives. We can discover the relevant conditions by testing hypotheses. The testing is achieved by examining the relationship between variables by varying them and tracking the consequences of doing so. The

⁹ Sorensen [1992a], 205.

¹⁰ Sorensen [1992a], 186.

relationship between the dependent and independent variables in an experiment can be expressed in a conditional sentence,

If <the independent variable(s)>, then <the dependent variable>.

The dependent variable is the health of the plant. The independent variables are chosen from those factors that we think may affect the health of the plant, factors such as sunlight, temperature, water and nourishment. In an experiment our goal is to find those independent variables which are relevant to the dependent variables, and also to discover how these independent variables are relevant. This is partly achieved by applying inductive knowledge from a previously acquired body of evidence. Our body of evidence consists of statements about previous observations which have been confirmed, and is expressed by generalized statements that form part of the theoretical context of the experiment.

Without a developed theoretical context, it is difficult to choose the independent variables. For example, in the plant experiment, without a developed theoretical context we might test independent variables that are extraneous. We might test whether the plant thrives in the presence of Picasso paintings. Referring to our present theoretical context, however, we know that while light affects the plant's health, other visual stimuli make no difference to it. Sometimes the independent variable is a group of propositions describing a situation from which the dependent variable obtains. Further thought experiments refine which aspects of the situation are, or are not, determinative.

Hacking suggests that we follow the advice of the physicist George Darwin and sometimes engage in experiments that are capricious. Darwin's example is blowing the trumpet to the tulips every morning for a month. Hacking suggests that "probably nothing would happen, but if something did happen, that would be a stupendous

discovery."¹¹ This approach has the virtue of allowing the imagination to work in an unfettered manner.

Suppose we are inspired by Darwin's exhortation and decide to examine whether our plant will thrive when Verdi arias are present in its aural environment. From previously acquired knowledge which makes up the theoretical context we know that this type of plant thrives if it is watered regularly, kept at room temperature in moderate sunlight and given a nutritional supplement once a month. Using this knowledge, we can set the conditions of the experiment.

The conditional sentences for the variables in this experiment are the following:

If the aural environment consists mainly of Verdi arias, then the plant thrives.

If the aural environment consists mainly of silence, then the plant thrives.

If the experiment confirms both sentences, then we are likely to suspect that the independent variable is not relevant to the dependent variable. That is, Verdi arias, in particular, and the quality of the aural environment, in general, do not affect the well-being of the plant. We may, however, decide to test further. For example, we may decide to test whether this species of plant thrives on Purcell. If Purcell makes no appreciable difference to the plant's health we may decide to terminate the experiment. We have not found that there is no aural environment which contributes positively to the health of the plant, but we have inductive reasons for supposing that the plant's health is not contingent on its aural environment.

However, imagine that we continue experimenting and find that the plant thrives partly as a result of an aural environment of rap music. We then refine the independent variable further until we discover what aspect of rap music is beneficial to the plant. We may try to isolate the predominate rap rhythm, volume or frequency. For example, if we

¹¹ Hacking [1983], 154.

think that it is the volume of noise, and not the particular noise, that affects the plant's health, we will then vary the volume until we find what level of volume affects the plant. Then we may try different sounds at the same volume.

Sometimes the result of an experiment may contradict a previously accepted belief found in the theoretical context. In this case, the theoretical context must be modified. The modification usually occurs after the experiment that led to the new belief is examined for error.

The Michelson-Morley experiments led to the acceptance of the idea that the speed of light is constant.¹² The experiments also led to the rejection of the idea of an æther. Previous to these experiments it was believed that light travels as a wave and that it must travel through some kind of propagating medium. If this were so, then the measured speed of light would differ depending on whether it is measured in the direction of the earth's spin or against the direction of the earth's spin. The Michelson-Morley experiments showed that if light travels with or against the direction of the earth's spin, the speed of light is constant.

Hacking's interesting discussion of the Michelson-Morley experiments shows how experiments develop as they progress.¹³ In this case, the question under examination changed as the theoretical context was modified. At first, the experiment was designed to question a theory about the æther that explained the relation between the actual position of the stars and our perception of their position. It was thought that the experiment refuted the idea that the earth moved in relation to the æther. The final result of the experiment was a revision of the theoretical context. The belief that there was a propagating medium for light, the æther, was subtracted from the background theory, and new ideas about light were sought and added.

¹² For a detailed account of the Michelson-Morley experiments, see Swenson [1972].

¹³ Hacking [1983], 159-161.

Hacking uses this experiment to illustrate how the interpretation of experimental data does not follow a linear progression to a single conclusion, an impression sometimes given by text-book style reporting of experiments.¹⁴ At the start, the experimenters were motivated to design a device that could measure the movement of the earth relative to the æther. In the end, the experiment added to the primary theory, in the case of theories about the speed of light, æther and the positions of the stars, while at the same time having repercussions for the background theories in terms of the possibilities for measurement. Michelson won a Nobel prize for the developments in the possibilities of measurement that the experiment instigated.

With the hindsight of history, the Michelson-Morley experiment can be described as an experiment testing a conditional sentence such as:

If there is no difference between the measured speed of light traveling with the earth's spin and traveling against the earth's spin, then the speed of light is constant and there is no such thing as the æther.

However, the experimenters were motivated by a very different objective. The conditional sentence which they wanted to test might be summarized:

If an apparatus (the interferometer) can be built to measure the motion of earth relative to the æther, then Maxwell's prediction (that it is impossible to achieve such a measurement) is wrong.

Hacking's examination of the Michelson-Morley experiments shows that it is not easy to classify the experimental process. Despite this, we can say that an experiment explores a theoretical context with the hope that it will have repercussions for that context

¹⁴ Hacking [1983], 174.

in that it will require us to add to, subtract from, or revise beliefs in the primary theory or background theories of the experiment.

In the Michelson-Morley experiment we have an example of an experimental result that both contradicted a belief within, and added a belief to, its primary theory. It also affected the background theories.

Consider now an observation that confirmed a theory. In 1965, background radiation of 3 degrees K, uniformly distributed through space, was observed by radioastronomers Arno Penzias and R.W. Wilson. The background radiation puzzled them, but at the same time a group of theoreticians from Princeton postulated that if there had been a Big Bang, such radiation would exist. The observations of Penzias and Wilson confirmed a set of statements in the theoretical context of astronomy.¹⁵ After recording the observation, Penzias and Wilson conducted further experiments to understand what might cause the background radiation. With work, they were able to eliminate potential sources as varied as roosting pigeons and instrumental errors. The reading indicating background radiation persisted. In this way, they clarified that their reading indicated uniformly distributed background radiation rather than experimental error.

Here, we have an example of observation and theory working together. In the plant experiment, there is an intention to pursue a particular question in relation to a particular theory. In the Michelson-Morley experiments, the motivating question led to unanticipated results for both the primary theory and the background theories. In the Penzias and Wilson experiment, an observation was made that fortuitously fitted with theory.

¹⁵ Wilson [1979], 113-134.

In his *System of Logic*, Mill examines methods to help the experimenter determine the relationship between independent and dependent variables. In his chapter entitled "Of the Four Methods of Experimental Inquiry" Mill gives methods of ascertaining causes in the experimental process. However, contrary to the title, this chapter covers five, and not four, methods of experimental inquiry. They are the method of agreement, the method of difference, the joint method of agreement and difference, the method of residues and the method of concomitant variations.¹⁶ We need not address these methods in detail here. However, in order to see how Mill's methods relate to the view of experiments given here, we can consider them in relation to the following well-known early thought experiment in ethics, *Gyges' Ring*, from the second book of *The Republic*:

But as for the second point, that those who practice it do so unwillingly and from want of power to commit injustice, we shall be most likely to apprehend that if we entertain some such supposition as this in thought — if we grant both to the just and the unjust license and power to do whatever they please, and then accompany them in imagination and see whither desire will conduct them. We should then catch the just man in the very act of resorting to the same conduct as the unjust man because of the self-advantage which every creature by its nature pursues as a good, while by the convention of law it is forcibly diverted to paying honor to 'equality'. The license that I mean would be most nearly such as would result from supposing them to have the power which men say once came to the ancestor of Gyges the Lydian. They relate that he was a shepherd in the service of the ruler at that time of Lydia, and that after a great deluge of rain and an earthquake the ground opened and a chasm appeared in the place where he was pasturing, and they say that he saw and wondered and went down into the chasm. And the story goes that he beheld other marvels there and a hollow bronze horse with little doors, and that he peeped in and saw a corpse within, as it seemed, of more than mortal stature, and that there was nothing else but a gold ring on its hand, which he took off, and so went forth. And when the shepherds held their customary assembly to make their monthly report to the king about the flocks, he also attended, wearing the ring. So as he sat there it chanced that he turned the collet of the ring toward himself, toward the inner part of his hand, and when this took place they say that he became invisible to those who sat by him and they spoke of him as absent, and that he was amazed, and again fumbling with the ring turned the collet outward and so became visible. On noting this he experimented with the ring to see if it possessed this virtue, and he found the result to be that when he turned the collet inward he became invisible, and when outward visible, and becoming aware of this, he immediately managed things so that he became

¹⁶ Mill [1843], 388-407.

one of the messengers who went up to the king, and on coming there he seduced the king's wife and with her aid set upon the king and slew him and possessed his kingdom.

If now there should be two such rings, and the just man should put on one and the unjust the other, no one could be found, it would seem, of such adamant temper as to persevere in justice and endure to refrain his hands from the possessions of others and not touch them, though he might with impunity take what he wished even from the market place, and enter houses and live with whom he pleased and slay and loose from bonds whomsoever he would, and in all other things conduct himself among mankind as the equal of a god. And in so acting he would do no differently from the other man, but both would pursue the same course. And yet this is a great proof, one might argue, that no one is just of his own will but only from constraint, in the belief that justice is not his personal good, inasmuch as every man, when he supposes himself to have the power to do wrong, does wrong. For that there is far more profit for him personally in injustice than in justice is what every man believes, and believes truly, as the proponent of this theory will maintain. For if anyone who had got such a license within his grasp should refuse to do any wrong or lay his hands on others' possessions, he would be regarded as most pitiable and a great fool by all who took note of it, though they would praise him before one another's faces, deceiving one another because of the fear of suffering injustice.¹⁷

Glaucon uses this thought experiment to demonstrate that justice is not chosen for its own sake, as Socrates claims, but rather for the sake of its consequences. Socrates is challenged to show that this is not so. Glaucon postulates that it is always more advantageous to be unjust, excepting when the societal reprisals one might face from unjust behaviour outweigh its advantages. He tests his hypothesis with the thought experiment in which the just and the unjust man are in the same situation — a situation in which they can act without it being possible to attribute their actions to them. In Glaucon's thought experiment, both the unjust and putatively just man act in the same way. They take advantage of their ability to be invisible in order to commit injustices and reap the benefits from them. The relationship between the variables in Glaucon's thought experiment can be seen in the following conditional sentences:

¹⁷ Plato [1930], 607-608.

If a just man can act without being seen, then he will act unjustly.

If an unjust man can act without being seen, then he will act unjustly.

Glaucon's thought experiment purports to confirm both conditionals. If we think Glaucon's scenario is plausible, the result of the experiment has repercussions for the theory given by Socrates in Book I of *The Republic*. It suggests that it is not always better (in the sense of most advantageous to the agent) to be just; in fact, in a situation in which it is possible to be invisible, it is better to be unjust. Further, it suggests that it is not wisdom that makes people act justly, but rather fear of reprisal.

We can use Mill's methods as a way to determine sufficient conditions for the dependent variable. We have seen that the question posed by the thought experiment can be expressed by conditional sentences which the experiment either confirms or disconfirms. The conditionals show the relationship between the dependent and independent variables. Glaucon's thought experiment can be characterized by the following table:

Independent variables (sufficient conditions)		Dependent variable (outcome)
Moral wisdom	Reprisals	Just action
Present	Absent	Absent
Absent	Absent	Absent

A sufficient condition for an outcome cannot be present when the outcome is absent. Here, it is the prospect of lack of reprisals that qualifies as a sufficient condition for unjust action, not lack of moral wisdom as Socrates claims. This supports Glaucon's position that people are not just due to their natures, but rather due to fear of reprisal.

Note that, strictly speaking, Glaucon has not proven anything with his thought experiment. He has given a convincing story. It is convincing because it agrees with our background theories about human sociology and psychology. Our responses to thought experiments are based on our intuitions which are supported by our knowledge of the theoretical context. In this case, the knowledge of the background theories of psychology and sociology affect our response to the thought experiment. From this knowledge we accept Glaucon's thought experiment as plausible.

Now, consider another thought experiment. This thought experiment, about a surgeon and his five seriously ill patients, is a variant of a type of thought experiment often constructed to argue against utilitarianism:

A surgeon has five patients who will die unless they are provided with certain essential body parts. A young man has just come in for his yearly check-up, and his parts will do: the surgeon can cut him up and transplant his parts among the five who need them. The surgeon asks the young man if he is willing to volunteer his parts and thus his life; the young man says "Sorry, I deeply sympathize with your five patients, but no." Would it be morally permissible for the surgeon to proceed anyway?¹⁸

Our response to the thought experiment is that it presents a challenge to utilitarianism. That is, in the situation described, we ought not to act in such a way to maximize utility. The result of this thought experiment has repercussions for utilitarian theory. The fact that such a case seems intuitively wrong leads us to consider rejecting or modifying utilitarianism. A rule utilitarian avoids the kind of problem that this thought experiment raises. A rule utilitarian acts according to the rule which, if adopted, leads to the most utility, rather than acting to maximize utility in each situation. In response to this thought experiment, the rule utilitarian would be likely to argue that there would be a rule forbidding the surgeon to use the young man's body parts, because if the contrary rule were adopted there would be fear in the population, and overall utility would not be enhanced.

¹⁸ Thomson [1986], 257.

In the course of examining the above examples, we have seen that the following statements apply to both experiments and thought experiments:

1. An experiment tests a statement (or set of statements) by varying variables relative to a theoretical context. This statement may be one that is a candidate for admission into the theoretical context or it may be a statement already in the theoretical context.
2. The more developed the theoretical context, the easier it is to set the experimental conditions and narrow the range of the independent variable.
3. The result of the experiment has repercussions for the theoretical context, which includes the primary theory, or set of theories, and the background theories of the thought experiment. It may add to, subtract from, or revise statements in the theoretical context.
4. Subsequent experiments are more refined in relation to the more refined theoretical context.

2.2 Thought Experiments and Arguments

There is disagreement both over what thought experiments *are* and what thought experiments *do*. Thought experiments have been variously described as "intuition pumps,"¹⁹ "picturesque arguments,"²⁰ and "windows to the world of the Platonic Forms."²¹ Norton is the most prominent proponent of the view that a thought experiment is an argument. Although the thought experiment may not be given as a carefully constructed argument, it can always be reconstructed as an argument.²² The alternative to this idea, Norton notes, might be something like Brown's idea that thought experiments provide a window from which to comprehend Platonic universal laws.²³

¹⁹ Dennett [1987], 323.

²⁰ Norton [1996], 334.

²¹ Brown [1991a], 127.

²² Norton [1991], 129.

²³ Brown [1991a], 127.

I agree with Norton's claim that thought experiments can always be reconstructed as arguments. However, I do not hold that all thought experiments are merely arguments. There is a distinction to be made between the manner in which a thought experiment may be characterized and the process of conducting a thought experiment. This distinction is similar to Reichenbach's distinction between the context of discovery and the context of justification.²⁴ On Reichenbach's account, in the process of justifying a proposition, any undesirable influences that may have played a part in the actual process of coming upon it may become apparent and taken into account. The process of thinking of a proposition or a theory is separate from the process of critical reconstruction. In the process of critical reconstruction the thought process is examined and reconstructed in a way that shows the justifiable connections between the propositions involved. As an example, Reichenbach notes the distinction between the way in which mathematicians or physicists conceive and develop theorems or theories and the way in which they present them to the public.²⁵

Reichenbach distinguishes the descriptive task of epistemology from the critical task.²⁶ The descriptive task, which is given in the rational reconstruction of the thinking, brings us part way in the critical process, since the reconstruction is conducted with the goal of producing a reconstruction which is justifiable. However, the rational reconstruction is bound to the original thinking by correspondence to that thinking. The critical task goes beyond that of the descriptive one. The goal of the critical task is to analyze the rational reconstruction to see whether justifiable connections between propositions do indeed exist.²⁷

In a similar way, reconstructing a thought experiment as an argument begins a process of justification operating on a less restricted process of discovery. In the critical

²⁴ Reichenbach [1938], 6f.

²⁵ Reichenbach [1938], 6.

²⁶ Reichenbach [1938], 7.

²⁷ Reichenbach [1938], 8.

process of justification, one can determine whether the argument, and thus the thought experiment, is a successful one. Norton claims that the success of the thought experiment can be equated with the success of the argument. Thought experiments will fail in the same way that arguments fail, when there has been a false assumption or a fallacious inference.²⁸ In this dissertation I argue that not only must the thought experiment contain strong argumentation, but it must also be grounded in its theoretical context.

A thought experiment may be presented as the scenario that first occurred to the thought experimenter, or it may be presented in a more accessible and rational order characteristic of a rational reconstruction. Both these forms of presentation omit the critical analysis necessary to determine whether the thought experiment is a successful one. The critical task takes the rational reconstruction and analyzes it for justifiable connections between the premisses. In this chapter, I reconstruct various thought experiments as arguments and analyze their success. In chapter 3, I suggest that after the rational reconstruction of the thought experiment has been examined for its deductive validity or inductive strength, then the experimental set-up of the thought experiment must also be assessed for the degree to which it is grounded in the theoretical context of the thought experiment. In chapter 6, I argue that responses to thought experiments must also be grounded. All of these processes are part of the critical examination of the thought experiment that concerns me in this dissertation.

In the context of discovery, then, a thought experiment may be more or less than an argument; in the context of justification, it must be an argument. The original thought experiment may include many details that do not need to be included in the argument form of the thought experiment and at the same time leave out assumptions which need to be included in the argument form of the experiment. Details added to help imagine the thought experiment situation may be omitted in its argument form. By contrast,

²⁸ Norton [1996], 356-357.

simplifying assumptions used to help isolate variables in the thought experiment must be noted in the argument's premisses and carried through to the conclusion. Here, I agree with Norton's observation that even though the thought experiment may not seem to be an argument in the context of discovery, it must be one in order to be a thought experiment at all. The argument, however, is often disguised because it is abbreviated or contains suppressed premisses.²⁹

The process of discovery may be facilitated any number of ways. The thought experiment process itself is probably as diverse and as difficult to characterize as the creative process of any pursuit. Thought experiments are certainly experienced and seem to be created in a manner different from that of creating arguments.

In his study of experiment, Hacking writes,

What is scientific method? Is it the experimental method? The question is wrongly posed. Why should there be *the* method of science? There is not just one way to build a house, or even to grow tomatoes. We should not expect something as motley as the growth of knowledge to be strapped to one methodology.³⁰

We cannot hope to classify the different ways in which experiments come about, but we may examine the experiment afterward and see if it is successful or not. We have seen some evidence in the previous section of Hacking's observation that experiment and theory interact in various ways; sometimes an empirical observation sparks an experiment which is then accommodated by theory, sometimes a theory suggests an experiment and sometimes an invention leads to new developments in theory.³¹

The genesis of a thought experiment, then, may be difficult to characterize. Careful reasoning can produce thought experiments. At the same time, reflection on such diverse influences as a newspaper article, play, chance suggestion, dream, or anecdote

²⁹ Norton [1996], 354.

³⁰ Hacking [1983], 152.

³¹ Hacking [1983], 154 f.

might generate a thought scenario which promises interesting results. The scenario may then be refined to isolate a certain variable and explore a certain question. As the thought experiment is refined, more theory is used and more detail is added to make the situation as clear as possible. Then the thought experiment can be examined for its success and we can ask several more precise questions: Does the thought experiment result in new knowledge? Does it give rise to new theory and new questions for observations? Can it be characterized as an argument? Is it a strong or valid argument? And, finally, is it grounded?

Recall physicist George Darwin's advice that we sometimes engage in experiments that are capricious.³² In thought, capricious experiments also can "work" but in retrospect, the ones that work are the ones that have the property of containing strong arguments that are grounded. The question of grounding can be settled retrospectively, which allows, if need be, the imagination to work in an unfettered manner. A thought experiment can also be carefully planned so that it is grounded before it is conducted.

Gooding claims that thought experimentation requires agency. He writes,

Thought experimental narratives show that agency is essential. Following such narratives requires participation in a process which is unintelligible unless we can recognize properties of the furniture of an imagined world well enough to understand procedures in it. The basis of such understanding lies in our personal knowledge: experience acquired through our own agency and its consequences. Such knowledge is presupposed in all of the narratives through which people have always learned about scientific experiment. As Davy remarked, situational, experiential knowledge may be tedious. Yet it is as essential to scientific thought as it is to everyday thought.³³

Gooding is addressing the process of discovery. In order to be able to create and follow thought experiments, we need to have certain experiential knowledge. I think the doing aspect of things has often been overlooked and it is important to see the interplay between

³² See §2.1.

³³ Gooding [1990], 215.

thought and action. In this thesis, however, I am concerned with the context of justification. I am concerned with how it is that we can examine a thought experiment and determine whether or not it is grounded.

Gooding makes an important observation about the division of experimental activity into the context of discovery and the context of justification. He suggests that the classification is not complete. Another important part of experiment is the conveying of the experimental procedure and results to a wider audience, which includes not only the rational reconstruction of the experiment, but the development of the experiment in responses to reactions from the audience. He notes that

[t]he development of field theory was shaped by work done in places such as workshops, lecture theaters, laboratories and demonstration rooms, and with ideas and images expressed in the literary contexts more familiar to historians, such as the pages of journals, an encyclopædia and manuscripts.³⁴

Gooding is writing about the experiments of Michael Faraday here, but thought experiments, too, or perhaps, particularly, develop in relation to the audience for which they were designed. Thought experiments are often modified as responses are sought and so develop over time. Searle's Chinese room, Putnam's twin earth and Thomson's ill-violinist thought experiments have all been modified and developed as they pass from thought experimenter to thought experimenter. In §7.3 we see how thought experiments can be modified and compared in order to calibrate the results.

On Gooding's account, experiments can be reconstructed in various ways and each way can be used to achieve a different end, be it exploration, communication, invention, consensus-seeking, argumentation or dissemination.³⁵ Here, I am interested in reconstruction of the logical argument form of the thought experiment. However, I agree with Gooding's observation that the distinction between context of discovery and context

³⁴ Gooding, Pinch and Schaffer [1989], 183.

³⁵ Gooding [1992], 49.

of justification oversimplifies the complex processes involved in the many different stages of an experiment.³⁶

There is a further distinction to be made. Thought experiments have two distinct parts. The first is the experimental set-up of the thought experiment; the second is the audience response to the thought experiment. When I say that a thought experiment can be reconstructed as an argument, the reconstruction includes both parts of the thought experiment. When I claim that a thought experiment must be grounded, both the experimental set-up of the thought experiment and the responses to the thought experiment must be grounded.

In sum, we have seen that in order to qualify as a thought experiment, the thought experiment must be amenable to reconstruction as an argument. Further, if there are putative thought experiments that cannot be later reconstructed as arguments, or sets of arguments, they are not thought experiments but examples of something else, perhaps some kind of description or intuition. We will see examples of non-thought experiments in §2.6.

2.3 Platonic Perception

I have claimed that thought experiments must be amenable to reconstruction as arguments. Now, recall Galileo's experiment about falling bodies given in Chapter 1. Here is the same thought experiment reconstructed as an argument:

Hypothesis:

When dropped, a heavier body does not fall faster than a lighter one provided both are of the same material.³⁷

³⁶ Gooding [1992], 49.

³⁷ The modern reader makes a simplifying assumption here, one which the characters in the dialogue are not in a position to make. We assume that the bodies fall in a vacuum. See my discussion of this point at the beginning of §2.4.2.

Premises:

1. Each falling body acquires a definite speed fixed by nature that cannot be increased or diminished except by the use of force. (from Aristotle)
2. Heavier bodies fall more quickly than lighter ones. (from Aristotle)

So,

3. When two bodies whose natural speeds are different are united, the faster (heavier) body will be slowed by the slower (lighter). (from 1 and 2)

and,

4. When two bodies whose natural speeds are different are united, the slower (lighter) body will be hastened by the faster (heavier). (from 1 and 2)

Subconclusion:

Therefore,

5. The united system will move more slowly than the heavier body alone and more quickly than the lighter body alone. (from 3 and 4)

Premiss:

6. The united system is heavier than the heavier body. (from experience)

Subconclusion:

Therefore,

7. The united system will move faster than the heavier body alone. (from 2 and 6)

Premiss:

8. Subconclusions 5 and 7 are contradictory.

Subconclusion

Therefore,

9. At least one of Premises 1, 2, 6, or 8 must be incorrect.

Assumption:

10. Premiss 2 is incorrect. Heavy and light objects fall at the same speed.³⁸

So,

11. The united system and each element of the system will fall at the same speed.

(from 10)

Conclusion:

Therefore,

12. Assumption 10 is more plausible than assumption 2, since it does not generate a contradiction; that is, heavy and light objects fall at the same speed.

Brown thinks that coming upon assumption 10 is a Platonic leap, a seeing into the laws of nature.³⁹ However, it is a standard process of investigation to modify a premiss that produces a contradiction. Premiss 2 can be modified to read "Heavier bodies fall more slowly than lighter ones", or "Heavy and light objects fall at the same speed". There are only two reasonable choices of alternate premisses for 2. The premiss "Heavy objects fall more slowly than light ones" is one that most people would not consider since it is easy to predict that the same contradiction would occur with this premiss as with the original one. When the experiment is run with each premiss in turn we find that a contradiction does not result when assumption 10 is employed. In fact, part of our background theory about theory selection includes the procedure of altering premisses in order to find which premiss or group of premisses leads to a non-contradictory state of affairs. There is nothing Platonic going on, just elementary logic.

³⁸ The reasons for choosing premiss 2 for modification rather than premisses 2, 6, or 8, are presented directly after the argument form of this thought experiment.

³⁹ Brown [1993b], 278.

We still need to consider why we choose to modify Premiss 2 rather than Premisses 1, 6, or 8. Each could be modified in a number of ways. Premiss 1 might be modified to state that, "Each falling body acquires a random speed fixed by nature that cannot be increased or decreased, except by force." Premiss 6 could read, "United systems have weights that are not affected by the weights of the objects of which they are composed." Finally, Premiss 8 could be modified so that it states, "Subconclusions 5 and 7 are both true." The choice of a particular premiss for modification and the way in which that premiss is modified depends on the theoretical context of the thought experiment. Every thought experiment has a primary theory, or set of theories, which is the theory, or set of theories, which directly bear on the question addressed by the thought experiment. In this case, the theory of physics is the primary theory. At the same time, there are background theories in operation. We choose to modify Premiss 2 because modifying it would cause the least disruption to our background beliefs. If we modify Premisses 1 and 6 we need to deny certain common knowledge about the constancy of physical laws for which we have inductive support from experience. If we choose to modify Premiss 8 we need to deny classical logic and accept a paraconsistent logic.

Quine's account of the web of belief elucidates the premiss modification choices in this experiment. The web of belief consists of observation sentences, non-observation sentences and laws of logic. When a theory is tested by a thought experiment and we are looking to revise it in the light of the results, we have a choice about which sentences to revise. That is, if, when we consider the argument:

P1

P2

P3

C

we find that $\sim C$, then we can conclude that $\sim (P1 \wedge P2 \wedge P3)$. This idea is given by the Duhem Thesis, which Duhem succinctly expresses in the following heading: "An Experiment in Physics Can Never Condemn an Isolated Hypothesis but Only a Whole Theoretical Group".⁴⁰ Quine restates this idea in his claim that "... theoretical sentences have their evidence not as single sentences but only as larger blocks of theory ...".⁴¹ How, then, do we choose which premiss to revise? We choose to reject the premiss that seems the least critical to the overall theory. Quine writes: "We heed a maxim of minimum mutilation."⁴² We can see why it is that laws of logic are the last to be revised. For these sentences are linked to all our information about the world and to reject any one of them would be to radically alter the web of belief of any theoretical context. Simplicity of the resulting theory is another guide in the modification of premisses. For example, a modification of premiss 1 that reads "Each falling body acquires a random speed fixed by nature that cannot be increased or decreased, except by force", would be rejected, since this premiss and the propositions one would have to add in order to accommodate it, would greatly complicate the web of belief. In altering the web of belief, then, we look to maximization of simplicity and minimization of mutilation.⁴³

The inference to the conclusion at 12 is another good example of the role which background theories play in thought experiments. The inference to 12 relies, not on the primary theory, but rather on a background theory about theory modification such as that described above. In our background theory about theory selection we choose to add those hypotheses or assumptions to our theories which do not cause inconsistencies. In this case, assumption 10, that all objects and combinations of objects fall at the same speed,

⁴⁰ Duhem [1906], 183.

⁴¹ Quine [1969], 80-81.

⁴² Quine [1990], 14.

⁴³ Quine [1990], 15.

does not lead to inconsistent states of affairs in which combinations of objects fall both more quickly and more slowly than one of the objects of the combination.

Brown supports the idea that there are Platonic thought experiments. He describes Platonic thought experiments as thought experiments that both refute a view and add a new view at the same time.⁴⁴ Brown notes that since these thought experiments add no new empirical data and are not logically deduced from old data, the explanation for their success is that somehow they allow us access to *a priori* knowledge of nature. Through Platonic thought experiments, Brown claims, we see the laws of nature themselves and not simply the regularities about which the laws are generalizations. Brown cites Galileo's thought experiment as an example of a Platonic thought experiment.

Brown compares the Platonic perception we have in Galileo's thought experiment with examples of Platonic perception of mathematical theorems. He claims that just as one can prove things with pictures in mathematics, so

there are thought experiments such as Galileo's which result in something like an immediate perception; they, too, are not arguments, but instead are vehicles for directing our attention so that we can simply see for ourselves — and I do mean *see*.⁴⁵

Consider the following theorem and diagram given by Brown:⁴⁶

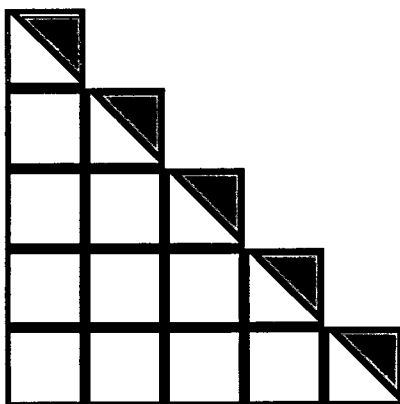
Theorem: $1 + 2 + 3 + \dots + n = \frac{n^2}{2} + \frac{n}{2}$

⁴⁴ Brown [1991b], 124f.

⁴⁵ Brown [1993b], 278.

⁴⁶ Brown [1993b], 275.

Proof:

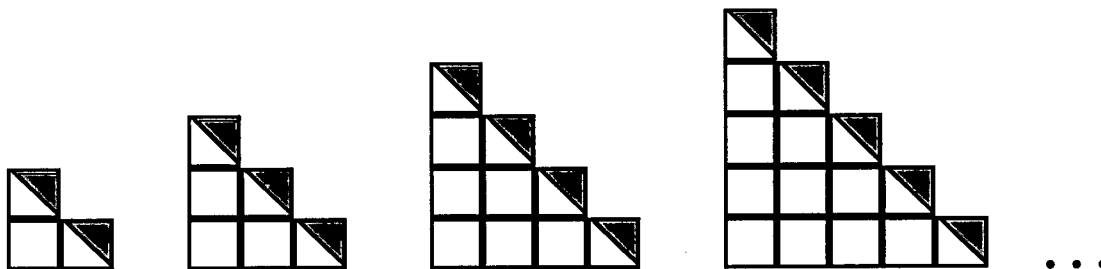


In order to prove a proposition to someone, one needs to provide sufficient reasons for the person to accept that particular proposition as true. This is achieved through linking previously accepted propositions by deductively sound or inductively strong operations to the proposition to be proved. In the case given above, Brown claims that the picture serves as a proof, because it allows us to instantly grasp the theorem and because upon grasping it, no further justification of the theorem is necessary.⁴⁷ This is not a traditional mathematical proof, he claims, but a Platonic one. On my account, the information gained by examining the picture needs reconstruction as an ordinary proof. We can build, step by step, from the given example(s) illustrated by the picture to the more general form expressed by the theorem.

Brown admits that interpretation is needed to make his picture a proof, and I agree. It is unlikely that there are instant perceptions derivable from diagrams. If one is quick, the perception may seem instant, but the quickness of one's thinking obscures the many steps that are needed, even in proving a relatively simple theorem such as the one that Brown presents. In order to appreciate the steps involved in the proof of this theorem, consider the possibility that a set of diagrams, such as the ones that follow,

⁴⁷ Brown [1993b], 276.

might come closer to a sort of visual proof of the theorem, much like a comic strip tells a kind of story in pictures:⁴⁸



When Brown says that his diagram is a proof I think we imagine a progression of diagrams such as those given above. Each step gives an instance of the theorem. The first diagram shows $1 + 2 = \frac{2^2}{2} + \frac{2}{2}$, the next shows $1 + 2 + 3 = \frac{3^2}{2} + \frac{3}{2}$, and so on.

However, even then, the pictures cannot give the proof. The inductive step needed to go from the instances to the theorem is not given by the picture.

Norton has given a verbal reconstruction of Brown's picture proof. This reconstruction underlines the fact that it is not Platonic perception but human reasoning that is at work here.

What has a staircase to do with arithmetic? Is this a joke? Brown is a funny man. Wait a moment. I think I see.

[1. Assumption] Each little square corresponds to an arithmetic unit.

[2. From figure] The figure consists of $(n = 5)$ columns of squares of height one, two, ..., $(n = 5)$

[3. From 1 and 2] so that the total number of small squares is the sum we seek, one + two + ... + $(n = 5)$.

But where is the theorem? There has to be some quick trick for getting it from the figure. Oh, I see it. It is easy.

[4. From figure] The total number of squares is the sum of the number of squares in the shaded and unshaded portions.

[5. From figure] The unshaded portion is half of an $(n = 5)$ by $(n = 5)$ square.

[6. Assumption] An $(n = 5)$ by $(n = 5)$ square has $(n = 5)^2$ unit squares in it.

[7. From 5 and 6] The unshaded portion has $\frac{(n = 5)^2}{2}$ unit squares.

[8. From figure] The shaded portion consists of one half square for each

⁴⁸ Hacking has a similar response to Brown's picture proofs. See Hacking [1993], 305.

column.

[9. From 2 and 8] There are $\frac{(n=5)}{2}$ shaded squares.

[10. From 4, 7, and 9] The total number of unit squares in $\frac{(n=5)^2}{2} + \frac{(n=5)}{2}$ which is the result sought.⁴⁹

Norton's reconstruction stops short of the inductive step that takes the instance of $1 + 2 + 3 + 4 + 5 = \frac{5^2}{2} + \frac{5}{2}$, and generalizes it to the given theorem. However, from his reconstruction we understand that the instant perception of diagram proofs, like coming upon assumption 10 in the Galileo thought experiment, is not instant at all. It is built on previous reasoning, experience and the resulting background theories which, having been often rehearsed, are easily accessed. The putatively instant perception, then, is built on past knowledge.

Gooding argues that the ability to visualize, along with other abilities we have by virtue of being embodied beings, is essential to thought experimentation. Rather than being a way to apprehend reality directly, thought experiments are a way to apprehend reality through our previously constructed representations of our embodied, knowing-how knowledge in the world.⁵⁰ This knowing-how knowledge is generalized from our input from our senses which we gather as we move through the world. For example, kinesthetic, visual, and aural cues help us to move and operate in the world and bring us information about the world. We use our reasoning to understand these interactions with the world. From this information we accept more generalized propositions which are available to us without the cue of the original stimuli. So, we understand because of many intermediary processes, some coming from being in the world, some coming from learning to reason. For example, Gooding writes:

⁴⁹ Norton [1996], 352-353.

⁵⁰ Gooding [1993], 285-286.

The experimenter can "see" directly, transparently, that (say) a collision involves a perfectly elastic recoil. Here many complex perceptual processes and theoretical judgments are distilled into one simple statement of the nature of the entities and their interactions. The experimenter *knows* that such is the case, without having to observe at all. Knowledge of a perfectly elastic recoil is not given through sensation: it combines ordinary experience of imperfectly-elastic objects with propositionally represented definitions.⁵¹

When Brown offers his account of picture proofs he glosses over the role that intermediary knowledge plays in perception, and it is this which gives the impression of Platonic insight.

2.4 Extraneous Particulars and Simplifying Assumptions

A thought experimenter devises an experimental set-up in order to elicit responses from the audience. The experimental set-up is comprised of the experimental conditions and the variables. Experimental conditions are just those features of the experiment that are not the variables. These conditions are set by considering the theoretical context.

Thought experiments employ simplifying assumptions in order to allow the question posed by the experiment to be highlighted by isolating the variables more effectively. Extraneous details may also be supplied to help the experimenter imagine the situation. Care must be taken to consider how all experimental conditions affect the variables, either directly or indirectly.

2.4.1 Extraneous Particulars

Consider a number of different presentations of Galileo's well-known thought experiment about the speed of falling objects. Each presentation varies in the number of extraneous particulars given. Recall, from Chapter 1, my examination of the original thought experiment from Galileo's *Two New Sciences*. The dialogue form, the three characters' conversation, the expressions of puzzlement and agreement, the examples of having a

⁵¹ Gooding [1993], 285.

heavy load on one's shoulders and trying to strike a running man with a lance, are examples of extraneous details. These instances are aids to understanding the argument, but they are not necessary to the argument itself.

Galileo's thought experiment could be given as a thought experiment with no more extraneous particulars than those which a physical experiment would present.

Consider this description:

Let us now look at the finest example of a thought experiment ever: Galileo's wonderful argument in the *Discorsi* to show that all bodies, regardless of their weight, fall at the same speed. It begins by noting Aristotle's view that heavier bodies fall faster than light ones ($H > L$). We are then asked to imagine that a heavy cannon ball is attached to a light musket ball. What would happen if they were released together? Reasoning in the Aristotelian manner leads to an absurd conclusion. First, the light ball will slow up the heavy one (acting as a kind of drag), so the speed of the combined system would be slower than the speed of the heavy ball falling alone ($H > H+L$). On the other hand, the combined system is heavier than the heavy ball alone, so it should fall faster ($H+L > H$). We now have the absurd consequence that the heavy ball is both faster and slower than the even heavier combined system. Thus, the Aristotelian theory of falling bodies is destroyed.

But the question remains, "Which falls faster?" The right answer is now plain as day. The paradox is resolved by making them equal; they all fall at the same speed ($H=L=H+L$).⁵²

There are some extraneous particulars in this presentation of Galileo's thought experiment, but fewer than in the original. Now consider the argument reconstruction of Galileo's thought experiment given in §2.3. In this presentation there are no extraneous particulars. We see that the same thought experiment can be given with varying numbers of extraneous particulars.

2.4.2 Simplifying Assumptions

Thought experiments often contain simplifying assumptions. Here we see an advantage of thought experiments over physical experiments. In physical experiments we must take care to isolate the variables from the influence of other factors. In thought experiments

⁵² Brown [1991b], 122-123.

we can just stipulate away confounding factors.⁵³ However, care must be exercised in employing simplifying assumptions.

In the Galileo thought experiment, modern readers tend to make a simplifying assumption that the characters in the dialogue are not in a position to make. Norton notes that while the characters in the dialogue clearly refer to the objects falling in a medium, modern readers are able to dismiss the air resistance of the medium through which the objects fall, and so reach the conclusion applying to bodies falling through a vacuum. In effect, modern readers make a legitimate simplifying assumption. However, Salviati, Sagreto and Simplicio are not in the position to make this simplifying assumption since this thought experiment occurs in the midst of a broader discussion about the possibility of the existence of a vacuum. In fact, the characters of the dialogue have some discussion about the effects of shape and friction on the velocity of descent. The thought experiment is pushed through by Salviati who claims that his conclusion about falling bodies is more empirically accurate than that of Aristotle.⁵⁴

Now, consider the thought experiment known as Stevin's chain. In it, the problem is to determine how much force is needed to keep a ball on an inclined plane stationary.

The conclusion is that:

On inclined planes of equal heights weights act in inverse proportion of the lengths of the planes.⁵⁵

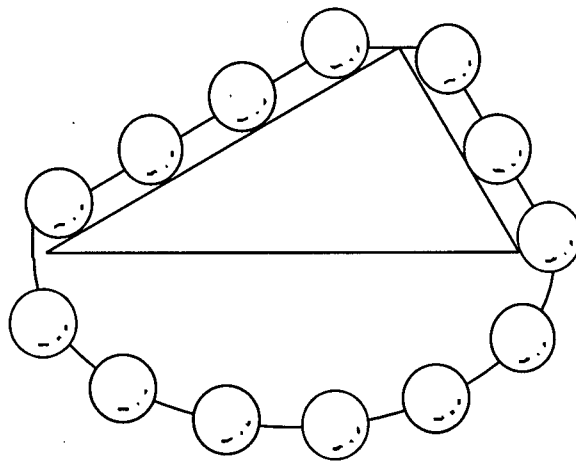
That is, the force needed to keep the ball in place on the inclined plane is inversely proportional to the length of the plane. To reach this conclusion, consider a chain draped over a prism, as in the diagram below:⁵⁶

⁵³ Sorensen [1992a], 205.

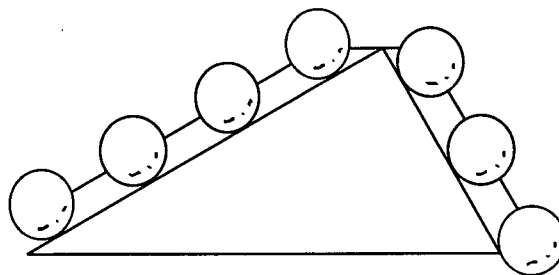
⁵⁴ Norton [1996], 343-345.

⁵⁵ Mach [1883], 34.

⁵⁶ The next two diagrams are modeled on those of Brown in [1993b], 271.



Next, apply the simplifying assumption that there is no friction. Then we imagine that we can set the chain in equilibrium. Mach claims that we instinctively know that this is possible.⁵⁷ I think that this is an example of Gooding's embodied knowledge, or "knowing-how." At this point, we see that when the chain is at rest, the part of the chain looped below the prism is superfluous, because the chain would be at rest with or without it. This is evident because the chain draped below the prism is symmetrical; it does not pull the remaining chain in either direction. If the chain hanging below the prism is detached, we get this diagram:



⁵⁷ Mach [1883], 34.

Then, when we consider that the weights of the chains on each side of the prism are proportional to their lengths, the result is Stevin's conclusion that "On inclined planes of equal heights, weights act in inverse proportion of the lengths of the planes."

In this experiment, we employ the simplifying assumption of frictionless surfaces. Consider the use of the simplifying assumption. This simplifying assumption allows the thought experimenter to isolate those variables pertaining to weight and length. If we had supposed that there were friction, we could not as easily have clarified the relationships between weight and length acting on inclined planes.

A simplifying assumption must be carried forward from the premisses of the thought experiment to the conclusion. This ensures that the conclusion is one that is related to the situation from which it arises. In this case, we conclude that in a frictionless situation Stevin's result holds, because we obtained the result from an example in which we stipulated away friction. Then, the effect of the simplifying assumption can be taken into account, in applications in the actual world. That is, we know that the effect of friction needs to be added to our calculations in situations in which Stevin's result is applied.

I have said that simplifying assumptions must be carried forward to the conclusion. At this point, they may fall out of the conclusion, if it turns out that they are irrelevant to the conclusion. For example, recall the ill-violinist thought experiment given in Chapter 1. In this thought experiment, Thomson uses a simplifying assumption to avoid the slippery slope problem of determining at what point of time a foetus becomes a person. Many anti-abortion arguments are based on the premiss that a foetus is a person from conception, or at some other defined stage, and so needs to be granted the rights of a person which include the right to life. Thomson assumes that the foetus has the status and

rights of a person from the beginning, thus avoiding the problem.⁵⁸ The result of the experiment — how we balance various rights — is not affected by this assumption. In this case, the simplifying assumption need not be carried forward to the conclusion. Through using a simplifying assumption, Thomson has separated issues about the personhood of foetuses from those of the rights of persons.

In Stevin's and Thomson's thought experiments we see two different results of simplifying assumptions. In both cases, the assumptions are used to isolate the variable. In Stevin's chain the simplifying assumption of frictionless surface is used to isolate the relationship of weight and length on inclined planes of equal height, and in Thomson's ill-violinist the simplifying assumption of the personhood of foetuses is used to isolate the relationship of the rights of persons to determine what happens in and to their bodies, and the rights of persons to life. In both cases, the simplifying assumptions are carried forward to the conclusion; however, in one case, the simplifying assumption (frictionless surfaces) remains in the conclusion, and in the other, the simplifying assumption (the personhood of foetuses) is irrelevant to the conclusion.

Laymon examines the use of counterfactual states of affairs in thought experiments and proposes that thought experiments can best be seen as ideal situations approaching real ones. He claims that both thought experiments and real experiments create situations that are idealizations of those in the world and then are made to fit with real events by theory-mediated correction. Thought experiments, he claims, need argumentation in order to show the relevance of the counterfactual situation to a real world situation.⁵⁹ The real experiment needs less theory-mediated correction since the experiment already happens within the world. The thought experimenter must be careful to understand how the thought experiment scenario approaches a real one.

⁵⁸ Thomson [1986], 38.

⁵⁹ Laymon [1991], 188.

Laymon's account suggests a distinction between types of thought experiments. Some use idealized real situations. We see this in Galileo's and Stevin's thought experiments. Others use fantastic but analogous situations. We see this in Thomson's ill-violinist thought experiment. Most thought experiments in science are idealized real situations, while thought experiments in ethics can be either analogous fantastic situations or idealized real situations. The thought experiment about the utilitarian surgeon and his five patients is an idealized real situation. What is important to note here is that idealized real situations must be made explicit by carrying idealized assumptions forward to the conclusion. Analogous fantastic situations, by contrast, have to be closely examined in order to see whether the analogy holds.

2.5 Logical Structure of Thought Experiments

Once a thought experiment is rendered into argument form, the type of reasoning employed becomes evident. Logically conclusive, or deductively valid, arguments are those in which, if the premisses are true, then the conclusion must also be true. By contrast, inconclusive arguments are those in which the premisses give evidence for the conclusion, but the conclusion does not follow from the premisses with a certainty of 1. The inductive strength of the argument is assessed by considering the degree of certainty with which the conclusion follows from the premisses. When we assess thought experiments, we assess them for deductive validity or inductive strength.

2.5.1 Deductive Thought Experiments

The deductive thought experiment is almost always a *reductio ad absurdum*. A *reductio ad absurdum* works this way:

You want to show that p .

Assume that not- p .

Show that not- p (together with accepted background conditions) leads validly to a

contradiction, or leads to something which is obviously false and agreed to be false.

Conclude that not- p is false since it leads validly to a false conclusion.

Conclude that p is true.

As an example of a deductive thought experiment, consider Euclid's proof of the existence of an infinity of primes:

Prime numbers are more than any assigned multitude of prime numbers.

Let A, B, C be the assigned prime numbers;

I say that there are more prime numbers than A, B, C.

For let the least number measured by A, B, C be taken, and let it be DE;

let the unit DF be added to DE.

Then EF is either prime or not.

First, let it be prime;

then the prime numbers A, B, C, EF have been found which are more than A, B, C.

Next, let EF not be prime;

therefore it is measured by some prime number,

Let it be measured by the prime number G.

I say that G is not the same with any of the numbers A, B, C.

For, if possible, let it be so.

Now A, B, C measure DE;

therefore G also will measure DE.

But it also measures EF,

Therefore G, being a number, will measure the remainder, the unit DF:

which is absurd.

Therefore, G is not the same with any one of the numbers A, B, C,

And by hypothesis it is prime.

Therefore the prime numbers A, B, C, G have been found which are more than the assigned multitude of A, B, C.⁶⁰

This is a straightforward *reductio ad absurdum*. Euclid assumes that there is not an infinity of primes and then shows that the assumption leads to a contradiction. He concludes that there is an infinity of primes.

The Galileo thought experiment about falling objects is also a *reductio ad absurdum*. In this case, Galileo demonstrates that objects of different masses fall at the

⁶⁰ Euclid, Book IX, Proposition 20.

same speed by showing that the negation of this claim leads to a contradiction. The contradiction is that the combined system both falls more quickly and more slowly than the smaller object.

The Thomson ill-violinist thought experiment is both an argument from analogy and a *reductio ad absurdum*. Consider Thomson's thought experiment, in argument form:

Hypothesis:

The right to life does not always outweigh the right to determine what happens in or to one's body.

Premises:

1. All persons have a right to life.
2. All persons have a right to decide what happens in and to their bodies.
3. One person's right to life outweighs another person's right to decide what happens in or to his or her body.
4. If you are medically connected to an ill violinist against your will and you disconnect yourself, you kill him.
5. If you are pregnant as a result of rape and you abort the foetus you kill it.
6. The violinist is a person.
7. The foetus is a person (from the moment of conception).
8. Being kidnapped and medically connected to an ill violinist for 9 months against one's will is analogous to being pregnant as a result of rape.
9. The application of moral principles and moral judgements should be consistent across analogous cases.

Subconclusion:

Therefore,

10. The violinist may not be disconnected. (from 1, 2, 3, 4, 6)

and

11. The foetus may not be aborted. (from 1, 2, 3, 5, 7)

Premisses:

12. Subconclusion 10 is unacceptable.⁶¹

Subconclusion:

13. If 10 is wrong, then so is 11. (from 8,9)

Premiss:

14. Of the premisses which lead to the unacceptable result (i.e., Subconclusion 10) it is Premiss 3 which is the least plausible.⁶²

Conclusion:

Therefore:

15. One person's right to life does not always outweigh another person's right to decide what happens in and to his or her body. (from 14)

In this *reductio ad absurdum* Thomson wants to show that a person's right to life does not always override a person's right to decide what happens in or to her body. She assumes the opposite in Premiss 3 and arrives at an unacceptable conclusion.

At first, one may consider that the *reductio ad absurdum* in the Thomson thought experiment does not operate in quite the same way as in the Euclid or Galileo thought experiment. For, instead of generating a contradiction from an assumption, Thomson's

⁶¹ This response is based on background considerations. See the discussion following this argument form.

⁶² See my account of premiss modification and plausibility in § 2.3.

reductio ad absurdum seems to generate an unacceptable result. However, on further examination we see that this thought experiment does operate in the same manner as any other *reductio ad absurdum*. The unacceptable result is unacceptable because it contradicts something in the theoretical context of the thought experiment. In this case, not unplugging the violinist is unacceptable because we have accepted the idea that we should not have to use our life to support another life without our consent. For example, we do not force people to be organ donors. This idea arises from our framework of values. In Chapter 4, we will see the role of the framework of values as the primary theory against which thought experiments concerning ethics are conducted. Thomson's thought experiment shows that we may initially accept Premiss 3 in the situation of the pregnant mother and not in the situation of the ill violinist. Since the situations are analogous and we wish to be consistent we should hold the negation of Premiss 3 in both cases.

Gooding holds that thought experiments work by setting up situations in which impracticability is featured rather than logical contradiction.⁶³ He writes:

What's involved is the demonstration, using a paradox, of an "impracticability" or contradiction in practice. Thought-experimenters criticize a rival theory by showing the impracticability of doing something in the way required *by that theory*. In short, a theory is criticized through the practice that links it to those aspects of the world that it purports to be about.⁶⁴

This idea follows from Gooding's thesis that thinking and doing are interrelated, even in thought experiments, and so what a thought experiment produces is not simply a logical contradiction but an impracticability. The impracticability, when set against the theoretical context and background assumptions, can be expressed as a contradiction. This point brings one back to the distinction between context of discovery and context of

⁶³ Gooding [1990], 204.

⁶⁴ Gooding [1990], 206.

justification. What is an impracticability in the context of discovery, becomes a contradiction in the context of justification. Consider again the ill violinist. What seems an impracticability here, that the kidnapped person should remain with the ill violinist,⁶⁵ becomes a logical contradiction in the argument form of the thought experiment. Gooding also makes this distinction. In discussing Galileo's thought experiment about an inclined plane⁶⁶ in which two incompatible criteria of speed are considered he writes:

The experiment works by showing that there is no consistent way of applying the criteria of speed then in use. "Consistency" here is a practical matter, that is, a matter of practicability. When verbalized it also becomes a conceptual matter and eventually, through thought experimentation, a logical one. But the force of the experiment as an argument involves a subtle mix of material- and mental-world manipulations.⁶⁷

Some thought experiments demonstrate impracticability or unacceptability as Gooding describes. Others, such as the Euclid thought experiment, give a logical contradiction. All of these thought experiments can be reconstructed as arguments in which the impracticability or unacceptability becomes a logical contradiction. Gooding is not far from the mark, then, when he states that "the formal consistency of theory and observation is an idealization of the practical consistency of the thinkable and the do-able."⁶⁸

2.5.2 Arguments by Analogy

Argument by analogy is often used in the thought experiment process. When we argue from analogy we employ the following type of reasoning:

⁶⁵ Note that Gooding is not addressing ethical situations which seem impractical, but things that won't work, *practically*, in the world of action. For Gooding, "impracticability" means "contradiction in practice." Gooding [1990], 206.

⁶⁶ I do not examine this thought experiment in this dissertation.

⁶⁷ Gooding [1990], 208.

⁶⁸ Gooding [1990], 209.

If situation x is like situation y , in that the independent and dependent variables in situation x are also applicable to situation y , and there are no confounding factors; then it is reasonable to expect that the conclusion we reach about the relationship between the variables in situation x is the same as the relationship between the variables in situation y .

In the Thomson thought experiment, Thomson argues that the situation of being kidnapped and hooked up to an ill violinist is similar to that of being pregnant due to rape. We have seen that prior to the thought experiment Thomson has granted, for the sake of argument, that a foetus has the status and rights of a person from the moment of conception. Thomson then argues that it does not follow from the granting of this assumption that abortion is morally impermissible. The thought experiment is designed to examine the relative weighting of the right to life and the right to decide what happens in and to one's body.

I have claimed that arguments using analogy only work if the situation described in the thought experiment is analogous to the other situation with respect to the variables in question, and there are no confounding variables. To test the former condition, we must see if it is possible to write a conditional sentence that applies equally to both situations. In the Thomson thought experiment such a sentence is available. It is this:

If another person obtains the use of your body for 9 months without your permission, and terminating their use of your body would result in death for that person, then it is (not) morally permissible for you to disconnect yourself from that person.

Since this sentence applies to both the situation described in the thought experiment and to the situation for which the thought experiment is analogous, we suspect that the analogy is a good one.

Next, we need to consider whether there are other factors which, when added to the situation, may override the analogy. Even if one can find a conditional sentence describing the dependent and independent variables which apply to both cases, the analogy may not hold for other reasons. Consider the following two cases:

The Shallow Pond. The path from the library at your university to the humanities lecture hall passes a shallow ornamental pond. On your way to give a lecture, you notice that a small child has fallen in and is in danger of drowning. If you wade in and pull the child out, it will mean getting your clothes muddy and either canceling your lecture or delaying it until you can find something clean and dry to wear. If you pass by the child, then, while you'll give your lecture on time, the child will die straight away. You pass by and, as expected, the child dies.⁶⁹

The Envelope. In your mailbox, there's something from (the US Committee for) UNICEF. After reading it through, you correctly believe that, unless you soon send in a check for \$100, then, instead of each living many more years, over thirty more children will die soon. But, you throw the material in your trash basket, including the convenient return envelope provided, you send nothing, and, instead of living many years, over thirty more children soon die than would have had you sent in the requested \$100.⁷⁰

The first case might be used to draw a distinction between harm by commission and harm by omission, and then the resulting principle may be said to hold for similar cases. The second thought experiment is given to show the parallel between the inaction in the case of the drowning child and the inaction in the case of the UNICEF donation.

The following conditional sentence applies to both situations:

If a child needs our help, and helping is easily within our power, then we ought to help.

However, despite the common conditional sentence, some may argue that the situations are not analogous for other reasons. Unger lists some such potential reasons: "physical proximity, social proximity, informative directness and experiential impact, thought of

⁶⁹ Singer [1972], quoted (with some paraphrasing) in Unger [1996], 9.

⁷⁰ Unger [1996], 9.

disastrous further future, unique potential saviours and multiple potential saviors, urgency".⁷¹ Unger considers and offers refutations of these many objections to the analogy between the shallow pond and the envelope thought experiments. On Unger's account, his thought experiments establish that the drowning child and the potential UNICEF recipients must all be saved.

Establishing that two situations are analogous is hard work. After establishing that the independent and dependent variables for the two situations are the same, careful consideration of what other factors of the two situations might challenge the analogy is in order.

Shelly Kagan's criticism of arguments by analogy is that they fall prey to two fallacious assumptions. The first is that

[t]he moral status of the act depends on whether the combined weight of the reasons for performing the act is greater or less than the combined weight of the reasons for not performing the act — and by how much.⁷²

This is the additive fallacy. At the same time, we assume that

if a factor has genuine moral relevance, then for any pair of cases, where the given factor varies while the others are held constant, the cases in that pair will differ in moral status.⁷³

Kagan calls this latter assumption "the ubiquity thesis". When we make the first assumption, the additive fallacy, then the ubiquity thesis makes sense, for the factor with moral relevance will always add to the sum in the same way, and changing any one factor will change the sum.

The reason that Kagan believes these to be fallacies is that there are cases in which a factor, which is usually morally relevant and bears weight on one's decision in an

⁷¹ Unger [1996], 33-48.

⁷² Kagan [1988], 14-15.

⁷³ Kagan [1988], 12.

additive manner, is not morally relevant. Assuming the ubiquity thesis and applying additive reasoning to these cases would lead to an incorrect result. Kagan gives the example of the killing and letting die distinction in the context of a situation where there is a motive for self-defense. He presents two cases:

In order to defend myself against the aggressor, I push him into a pit, expecting the fall to kill him;

In order to defend myself against the aggressor, I refrain from warning him about the pit into which he is about to fall, and I expect the fall to kill him.⁷⁴

On the ubiquity thesis, if a factor is relevant in one pair of cases, it is relevant in others. However, when the motive of self-defense is a factor in the situation, the killing-letting die distinction is not a factor in determining whether one action is better than another. If the relationship between factors is given by the equation: $S = x \cdot y + z$, then it may be that $x = 0$ in the case of self defense and 1 in the case of non-self defense. Then, if y is the killing-letting die distinction, it has no effect when self-defense is involved.⁷⁵ Kagan concludes that the additive model does not work in this case, and applying the killing-letting die distinction to this case on the grounds of the ubiquity thesis would yield a counter-intuitive result.

In contrast to the ubiquity thesis, Kagan suggests that a given factor will cause the same effect on analogous situations only when that factor plays an analogous role in those situations. For example, the role that oxygen plays in chemical reactions differs in particular cases. Sometimes it is determinative in giving the effect of combustion of a particular compound. In other cases it is not. The laws of the chemistry concerning oxygen are constant but the role of oxygen in each particular situation is not.⁷⁶

⁷⁴ Kagan [1988], 18.

⁷⁵ Kagan [1988], 19.

⁷⁶ Kagan [1988], 13.

Dealing with the over-application of the ubiquity thesis and the additive fallacy means ascertaining whether the roles of the considered factors are analogous and then whether there are any confounding variables that interact with the factors in a way that makes the situations disanalogous. Sorensen claims that the additive fallacy can be preempted by determining as closely as possible how variables interact. He writes:

The proper degree of concern can be ascertained by letting ordinary experimentation be our guide. Experimentalists have their share of cautionary tales about interacting variables. But these are the exceptional cases, ones that do not stop them from robust allegiance to Mill's method of difference.⁷⁷

Dennett raises the problem of oversimplified thought experiments which seem to be analogous to more complex situations but are not. He cites a thought experiment given by Hofstadter in which human behaviour is compared to that of wasps of the species known as the *sphex ichneumoeus*. The wasps' behaviour patterns seem to be deliberate but turn out to be mechanical responses to environmental conditions. Hofstadter suggests that since we differ only in degree from wasps, we may not be free agents. Dennett criticizes the over-simplification of the analogy. The analogy does not hold because there is a difference in complexity between wasps and humans and the causes of their respective behaviour. Dennett writes:

Might it not be that what makes the wasp's fate so dreadful is not that her actions and "decisions" are caused but precisely that they are so simply caused? If so, then acknowledged difference between the object of our intuition pump and ourselves — our complexity — may block our inheritance of the awfulness we see in the simple case.⁷⁸

Kagan calls these types of cases "transport cases", because they transport their conclusion from a simpler case to another more difficult case. We have seen that he criticizes them for adopting the ubiquity thesis and the additive fallacy.

Dancy also argues against transport cases. He writes:

⁷⁷ Sorensen [1992a], 273.

⁷⁸ Dennett [1984], 12.

The original idea was that we found the actual case difficult, and so we derived help from a case which was easier. So there is a further difference between the two cases; in one we find it hard to discern what we ought to do, and in the other we find it easy. But this could only be so if there were other relevant differences between the two cases, in virtue of which one is hard and the other easy. The one that is easy can only be so because it does not contain factors which complicate the issue in the hard case. And, this being so, it is odd to suppose that our decision in the easy case should be of help when we come to the actual hard case.⁷⁹

Simple cases that transport their conclusion back to complex cases do rest on the assumption that, *ceteris paribus*, factors make separate and independent contributions to situations, and that if a factor is morally significant in one situation it is so in all others. Despite this, simplified cases may be usefully employed as good analogies for more difficult cases and serve a purpose in helping to clarify responses to those difficult cases, if proper vigilance for the role of confounding variables is applied.

2.5.3 Inductive Thought Experiments

Arguing by analogy is a kind of ampliative or inconclusive reasoning. In this section, I examine another common type of ampliative or inconclusive reasoning used in thought experiments: induction by simple enumeration.

In induction by simple enumeration, a property of one object is discovered to hold of all observed objects of that kind. So we reason that the next, yet to be observed, object of that kind will have that property. In induction through analogy, as we have seen in §2.5.2, two analogous things or situations are said to share a feature that is observed in one case but not in the other.

In order to understand the inductive thought experiment, it is useful to examine Hume's and Popper's contrasting views on induction.

In *An Enquiry Concerning Human Understanding*, Hume poses the problem of induction with the following observation about the difference between two propositions:

⁷⁹ Dancy [1985], 146.

one describing a set of observations, and the other predicting future observations from this set of observations:

These two propositions are far from being the same: I have found that such an object has always been attended with such an effect, and I foresee that other objects which are in appearance similar will be attended with similar effects. I shall allow, if you please, that the one proposition may justly be inferred from the other: I know, in fact, that it always is inferred. But if you insist that the inference is made by a chain of reasoning, I desire you to produce that reasoning.⁸⁰

Hume felt that the practice of inferring from regularly occurring events in the past to the probable continuation of them in the future, while a common practice, is not grounded in reason. Hume presents the example of trees which in the past have been observed to flourish in May and June and decay in December and January. Hume asks if we can understand the proposition "that all the trees will flourish in December and January and decay in May and June." Since this is obviously an intelligible statement which "can be distinctly conceived, implies no contradiction, and can never be proved false by any demonstrative argument or abstract reasoning *a priori*," Hume asserts that it is not sound reasoning to "put trust in past experience and make it the standard of our future judgement."⁸¹ In other words, if we can conceive of something different happening from that which we have observed to happen regularly in the past, and if this idea implies no contradiction, then it is logically consistent to conceive that what we have previously observed may not continue in the same way.

Recall my earlier discussion of Reichenbach's distinction between the context of discovery and the context of justification.⁸² In the context of discovery we may notice that "such an object has always been attended with such an effect", and we may go on to reason that "other objects which are in appearance similar will be attended with similar

⁸⁰ Hume [1748], 48.

⁸¹ Hume [1748], 49.

⁸² § 2.2.

effects". However, in the rational reconstruction of this process of observing and reasoning, that is, in the context of justification, we may either find, as Hume does, that this process is not grounded in reason, or, we may find some way to justify it.

Hume claims that if we attempt to justify our use of induction we create a vicious circle, because our only possible proof of induction will be an inductive one. We can justify induction only by observing that inductive reasoning has been reliable in the past and so, by an inductive move, we conclude that it will be reliable in the future. However, a regress is created if we rely on past experience to justify our reliance on past experience to predict future events.

Hume challenges his reader to find a "necessary connexion" between two events that seem to be causally related. In asking for a "necessary connexion" Hume looks beyond spatio-temporal proximity, and memory of regularities in the past. He concludes that there is no such necessary connection, but rather habitual association of events which produce expectations that the association will continue. Hume writes:

Custom, then, is the great guide of human life. It is that principle alone which renders our experience useful to us, and makes us expect, for the future, a train of events similar to those of the past.⁸³

Popper views the problem of induction differently. In *The Logic Of Scientific Discovery*, Popper supports a scientific methodology that is based, not on induction, but rather on deduction. Popper writes that "not the verifiability but the falsifiability of a system is to be taken as a criterion of demarcation."⁸⁴ In other words, a scientific theory must be presented so that it can be singled out, not necessarily in a positive sense, but so that it can be singled out in a negative sense, by means of empirical tests.⁸⁵ Popper

⁸³ Hume [1748], 58.

⁸⁴ Popper [1959], 40.

⁸⁵ Popper [1959], 40-41.

points out that there is an "asymmetry between verifiability and falsifiability" that comes from the fact that universal statements are:

never derivable from singular statements but can be contradicted by singular statements. Consequently, it is possible by means of purely deductive inferences (with the help of the *modus tollens* of classical logic) to argue from the truth of singular statements to the falsity of universal statements.⁸⁶

The deductive inference to which Popper refers is:

$$H \supset O$$

$$\underline{\sim O}$$

$$\sim H;$$

where H is a hypothesis, and O stands for various observations that are predicted to occur if the hypothesis is correct.

By contrast, inductive methodology uses the following form of inference:

$$H \supset O$$

$$\underline{O}$$

$$H$$

Deductively, this is known as the fallacy of affirming the consequent, for it can be shown that using this method, given true premisses, one can still arrive at a false conclusion.

A problem with Popper's deductivism is evident when we consider Duhem's thesis, which I discussed in §2.3. Remember, that Duhem's thesis states that it is hypotheses and background assumptions together which imply predictions ($H \wedge B \supset O$). So when a prediction fails to obtain ($\sim O$), we can only know that we must reject either the hypothesis or the background assumptions ($\sim (H \wedge B)$). Consequently, by

⁸⁶ Popper [1959], 41.

modifying the background assumptions, the hypothesis can be saved from the evidence of observation. Popper's deductivism suggests that single hypotheses can be falsified.

Duhem's thesis makes it clear that it is not the single hypothesis that is falsified, but the whole of the theoretical context. Duhem writes:

In sum, the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one should be changed.⁸⁷

Two examples discussed by Gillies will serve to illustrate how the Duhem thesis highlights an important aspect of the process of modifying scientific theories which is not captured by Popper's deductivism. Consider the following two cases:⁸⁸

1. At the time when Uranus was the most distant planet known, astronomers noted that the observed orbit of Uranus did not agree with the theoretical orbit, calculated using Newton's theory and auxiliary hypotheses. This meant that either Newton's theory or one of the auxiliary hypotheses was false. Adams and Leverrier rejected the auxiliary hypothesis regarding the number of planets, and then predicted where a further planet would have to be, and what its mass would have to be, in order for it to cause the observed orbit of Uranus. On September, 23rd, 1948, the planet was observed very close to its predicted position.
2. Astronomers were also concerned about anomalies in the orbit of Mercury. Leverrier tried to modify the same auxiliary hypothesis as he had previously, the hypothesis about the number of planets, and he predicted the planet Vulcan, between the Sun and Mercury. No such planet could be found. It turned out that the anomaly in the orbit of Mercury was explained by the general theory of relativity, which replaced Newton's theory.

⁸⁷ Duhem [1906], 187.

⁸⁸ Gillies [1993], 100.

Gillies concludes:

We see that although the Uranus anomaly and the Mercury anomaly were *prima facie* very similar, success was obtained in one case by altering an auxiliary hypothesis, in the other by altering the main theory.⁸⁹

Popper holds a correspondence theory of truth. For Popper, scientific theories are tested piecemeal. When testing one hypothesis, the scientist must accept other background statements. We see in considering the Duhem thesis that there is a problem with this characterization of the scientific process. The coherentist view, in which a theory is a web of beliefs, the whole of which is tested by observation, avoids the problems that Popper's theory raises, and at the same time gives a satisfactory defense of induction. The coherentist justifies using induction if it leads to increased coherence in one's belief set. From this point of view, inductive practice is not solely a function of human nature as Hume claimed, but rather a reasonable practice that leads toward truth in the form of a more coherent belief set.

How is it that inductive principles of reasoning lead to increased coherence of one's beliefs set? If we say that we know this from past experience, we are engaged in a circularity. The coherentist response is that inductive inference is inference to the best explanation. The inductive inference provides the best explanation of all the evidence. And this of course is a more coherent explanation.⁹⁰ Consider how we can justify a particular belief about what will happen next. Dancy gives the example of a brick flying toward a window and our belief that the window will break.⁹¹ We believe that the window will break because there is a natural necessity that is basic to the possibility of explanation. We know that the window will break by inference. Dancy writes:

The passage of the brick entails the breaking of the window; for given the brick the window must break. And we know this because it would be far

⁸⁹ Gillies [1993], 100.

⁹⁰ Dancy [1985], 208.

⁹¹ Dancy [1985], 209.

harder to explain the window's not breaking than it would be to explain its breaking. (There are in fact links of mutual explanation.)⁹²

Induction is widely used in thought experiments. My focus in this dissertation is thought experiments concerning ethics. In Chapter 4, I present a contextualist account of ethics, and it is from this viewpoint that I examine the efficacy of thought experiments in ethics. In particular, a contextualist does not seek general conclusions about certain types of situations, but rather in-depth understanding of all the aspects of a particular situation. The contextualist approach is a non-foundational approach. That is, a problem is not resolved by considering an ethical rule, based on a more foundational ethical principle. Rather, the problem is resolved by considering how all the constraints that bear upon it can be balanced in a coherent manner. Given this, we face the concern of how, on the contextualist account, a thought experiment concerning ethics can be inductive. I find that the thought experiment methodology is excellent for using inductive results generated from cases that are similar while also assessing the particular features of the case at hand. Here is an example:

Consider, then, the case of a 74-year-old woman with a cardiac condition residing in a retirement home. An avid baseball fan who has many friends, she enjoys a very full life. Having witnessed what has happened to one or two of her acquaintances who have been resuscitated following cardiac arrest — brain-damaged and surviving on machines — she makes it clear to her doctor that she does not want to be resuscitated if she suffers such an arrest. Suppose, now, that she experiences a fairly mild arrhythmia and is consequently moved to the infirmary in order to try some different medications. Owing to an allergic reaction, a trial medication causes her heart to go into fibrillation and then full-scale arrest. We know that the adverse reaction to the drug can be reversed if she is resuscitated. More-over, because the arrest was caused by the allergic reaction, and because the code team is immediately available, we shall say that her chances of a full recovery, following resuscitation, are about 75-80 percent. The downside, of course, is the still significant chance of a permanently weakened heart and brain-damage — exactly what she most dreads.

The principle of autonomy bids us do what our patient would want for herself. But her desires have now become quite ambiguous. Did she

⁹² Dancy [1985], 209-210.

have *this* sort of thing in mind? Would not a rational person want resuscitation with these odds, especially in light of the quality of the life that she will most likely recover? Or does none of this matter? In short, how are we to morally interpret this situation?⁹³

This thought experiment describes a situation in which situational factors seem to outweigh standard ethical approaches. Of course, the principle that we respect the patient's wishes has a tacit *ceteris paribus* clause. In this case, we are to reassess the situation and imagine what the patient really would wish. We are working to assess the application of a principle in a situation in which the *ceteris paribus* conditions are not met. *Ceteris paribus* conditions help us with the difficulty of formulating principles that can be applied to every situation, but they leave us with principles that may not tell us much about any particular situation.

The results of this thought experiment will bear on a situation that is similar to this one. We use the thought experiment to understand and explore situational factors in as much detail as possible and then apply our ideas to other cases we deem similar. Note that the inductive move is not taken from a particular principle to all cases which fall under it. In this situation, for example, the principle of autonomy may not apply. Rather, the inductive move is from the particular response that this thought experiment suggests to the response to cases relevantly similar to it. However, this response is considered in relation with the other factors influencing the particular case at hand. This is the process of wide reflective equilibrium which is explained in detail in Chapter 4.

2.5.4 Thought Experiments which Test Explanatory Potential

Some thought experiments test the explanatory potential of the theory rather than the truth of the statements comprising the theory.⁹⁴ Thought experiments such as those of

⁹³ Winkler [1993], 355.

⁹⁴ Lennox [1991], 223-224.

Darwin in *On the Origin of Species* do just this. Darwin's thought experiments present the argument that

*if each of the mechanisms and processes referred to by Darwin's theory were to interact in particular ways, there would occur an accumulation of minute, random variations in a particular direction, culminating in distinct varieties and, eventually, new species.*⁹⁵

Consider this example:

In order to make it clear how, as I believe, natural selection acts, I must beg permission to give one or two imaginary illustrations. Let us take the case of a wolf, which preys on various animals, securing some by craft, some by strength, some by fleetness; and let us suppose that the fleetest prey, a deer for instance, had from any change in the country increased in numbers, or that other prey had decreased in numbers, during the season of the year when the wolf is hardest pressed for food. I can under such circumstances see no reason to doubt that the swiftest and slimmest wolves would have the best chance of surviving, and so be preserved or selected — provided always that they retained strength to master their prey at this or at some other period of the year, when they might be compelled to prey on other animals.⁹⁶

Thought experiments such as those of Darwin serve a different function from those that use straightforward induction or deduction. In other thought experiments we aim for a revision or a confirmation of an hypothesis within a theory. The revision takes the form of adding or revising statements in the primary theory or background theories of the thought experiment. In Darwin's thought experiments the statements of the theory itself are not tested. Instead, it is the explanatory potential of the theory which is at stake. The result is that the theory may be corroborated, discarded or modified. The thought experiments that Darwin presents are meant to explore the theory's explanatory capabilities. They are not meant to test whether the theory directly applies to the world.⁹⁷

⁹⁵ Lennox [1991], 229.

⁹⁶ Darwin [1859], 90.

⁹⁷ Lennox [1991], 241.

The type of reasoning that Darwin employs is evident in many fields of enquiry. Russell holds that mathematics operates by the principle of induction in the following passage:

[W]e tend to believe the premises because we can see that their consequences are true, instead of believing the consequences because we know the premises to be true. But the inferring of premises from consequences is the essence of induction; thus the method of investigating the principles of mathematics is really an inductive method, and is substantially the same as the method of discovering general laws in any other science.⁹⁸

A thought experiment provides one means to test the explanatory potential of a theory. If the thought experiment is successful, the explanatory potential of the theory needs to be evaluated on further grounds.

Explanations can be appropriate given the scientific theories and general background theories of the time, or with respect to individual epistemic states. A satisfactory characterization of explanation would give an account that applies to both cases. In general, since individual epistemic states are often in accord with background theories of the time, an account that works at one level will implicitly encompass the other. Two promising accounts of explanations are those of Kitcher and Gärdenfors. On Kitcher's account explanation is unification. A good explanation is that which accounts for the most phenomena included in one's belief set with the fewest and most stringent patterns of argument.⁹⁹ By contrast, Gärdenfors claims that the best explanation is the one which would diminish the surprise of the fact to be explained within the epistemic state of the agent who did not know the thing to be explained.¹⁰⁰

A successful thought experiment employs sound or strong reasoning and is grounded. In this dissertation, I am concerned with the grounding of thought

⁹⁸ Russell [1907], 273f. See Irvine [1989] for further discussion.

⁹⁹ Kitcher [1981], 339.

¹⁰⁰ Gärdenfors [1988], 177-178.

experiments. However, if the reasoning applied in the thought experiment is not valid or strong, the thought experiment, whether grounded or not, cannot succeed. For those thought experiments that test the explanatory potential of a theory we must evaluate the success of the explanation. In order to assess the success of an explanation, one must examine the explanation in relation to the theoretical context. Accounts such as those of Gardenförs and Kitcher need to be considered in evaluating these explanations. In §5.4, I examine the need to have grounded explanations and explain in more detail the relationship of a grounded explanation to its theoretical context.

2.5.5 Ampliative Reasoning in Computer Simulations

Sorensen writes that simulations "are thought experiments once removed."¹⁰¹ They don't directly manipulate variables in the situation that is of interest, but create another situation that is analogous to it. In this way a simulation is no different than any other thought experiment employing analogy such as Thomson's ill-violinist.

There are two main types of simulations: simulation demonstrations and simulation experiments. Simulation demonstrations are tools used to demonstrate a theory. An example of a simulation demonstration is a flight simulator. Flight simulators are used as tools to allow people to learn how to fly. The simulation is a demonstration rather than an experiment since it does not explore a question in relation to a theoretical context. The information about flying could be derived directly from theory, but the experience of flying could not. The simulation is a tool that provides an easier, and more interactive way, of demonstrating the basics of flying.

In this thesis, I am concerned with thought experiments and their cousins, simulation experiments. Simulation experiments test the explanatory potential of a theory, or propose causal links between one state of affairs and another. Consider

¹⁰¹ Sorensen [1992a], 225.

Schelling's models concerning segregation: Schelling showed with a simple spatially distributed model that when people prefer to avoid minority status, even if they are tolerant of living with those of other races, segregated neighbourhoods result. This is because no mixture except that which is 50:50 (assuming there are two races in question) will be self-sustaining. His models show that extreme segregation can occur even in the absence of extreme racism.¹⁰² Schelling's models were originally demonstrated without the use of computers; however, they are a good example of the spatially distributed models which computers execute well.

With computer simulations one can easily refine hypotheses. For example, in the Schelling case, one could test what degree of segregation occurs when various percentages of people have varying degrees of preference for neighbours of the same race. The ease of refining hypotheses and running simulations is a virtue of computer simulations. Computers can run more, and finer-grained, tests in a shorter period of time than can the human brain.

A good example of an imaginary situation analogous to Schelling's simulation is given by Resnick.¹⁰³ Imagine that there are two groups of about ten people at a cocktail party. Although each person might prefer a mixed gender grouping, each person is uncomfortable when their gender forms less than a third of the group. Imagine that one of the clusters ends up, by chance, with three men and seven women, and that one of the men leaves and joins another group. The remaining two men, feeling uncomfortable, drift away to the other group also. The result is a group of women, who will be joined by the women in the other group, as they feel their minority shrinking with the influx of men. This simple sketch shows how individual preferences for gender distribution can

¹⁰² Schelling [1978], 141f.

¹⁰³ Resnick [1994], 86-87.

lead to single gender conversation clusters, even if all people prefer mixed-gender clusters.

Using these simple examples we can examine the type of reasoning involved in simulations. Schelling wishes to consider the hypothesis that segregated neighbourhoods might result from a population whose members have some degree of preference for interaction with people of the same race over those of other races, but no objection to a certain amount of interaction with people of other races, combined with a certain initial mixture. A model is devised in which various levels of these preferences are tested and results recorded.

The model tests the hypothesis by trying to establish that the fact is consistent with the hypothesis. In this case, the facts are also consistent with the hypothesis that full-blown racist tendencies are in operation. It is interesting to consider the causal links in this example. We are looking for the cause of a known fact. In this case, we can only hypothesize about the cause. With simulations we can imagine situations and see if we can duplicate the effect, but we can never know whether our simulation is the only way of generating that effect. We can say we have evidence showing that our hypothesis is a likely hypothesis. To do this we would need to get evidence in favour of our hypothesis over any plausible alternative hypotheses. In this case, we would need to get evidence about the degree of racism that is responsible for segregated neighbourhoods.

Simulations are types of analogies. We claim that the situation simulated is relevantly similar to the situation we wish to study. All the considerations needed to assess arguments by analogy are needed here. We need to ask whether there are relevant differences between the simulation and the situation being simulated, and we need to watch for interacting variables. Simplifying assumptions need to be accounted for in the conclusion.

Now consider a computer simulation of the flocking behaviour of birds. In thinking about flocks of birds, most people assume that flocks have leaders, and that it is these leaders who initiate and terminate flight, and who set the direction and speed of the flock. Simulators of flocks have found that this explanation is likely to be false. Good simulations of flocking behaviour have been achieved without designating any leader. The flock is easily replicated by giving each individual computer generated "bird" a few simple rules. A good example of computer generated flocking can be seen in the film *Batman Returns*. The flock of bats flooding the underground tunnels, and the mob of marching penguins in the streets of Gotham city were simulated by creating a mob of identical bats, or penguins. The motion of each of the individuals in the group was then set to be governed by a few simple rules.

The flocking rules were discovered by Craig Reynolds, whose flocks consisted of computer generated entities guided by the following rules: do not bump into one another, keep up with the others, and do not stray too far away from one another.¹⁰⁴ The computer simulations are so realistic that biologists have re-examined films of flocking birds and decided that the flocking birds are following some simple local rules, which involve maintaining the flock by reacting to the birds nearby. Flocking behaviour is coordinated without a coordinator. Science writer James Gleick reports:

High-speed film [of flocks turning to avoid predators] reveals that the turning motion travels through the flock as a wave, passing from bird to bird in the space of about one-seventieth of a second. That is far less than the bird's reaction time.¹⁰⁵

The flocking behaviour seems to be explained by a number of simple rules. The rules are tested by computer simulation and the resultant flocking behaviour is achieved. In this case, we conclude that these rules can produce flocking behaviour. We accept the

¹⁰⁴ Kelly [1994], 11. See Reynolds [1987] for the original paper describing the flocking simulations.

¹⁰⁵ Kelly [1994], 10.

rules because they are consistent with what we know about the abilities of birds and because they do not conflict with other likely explanations.

In the course of simulating flocking behaviour the simulator must learn about the mechanics of flocking. The simulation may result in a modification of the theory of flocking. The rules used to run the simulation either produce a realistic analogue of a real flock, or they do not. If they do, they may induce additions or modifications to the theoretical context of the simulation, flocking theory. If they do not, the simulator must decide whether the lack of realistic simulation is a reason to discard, modify, or add to the rules that are used to govern the simulation, and what effect the lack of realistic simulation has on the overall theoretical context.

Now, consider the difference between this simulation and the segregation simulation. In the segregation simulation we see that both same race preference and extreme racism can bring about segregated neighbourhoods. However, we need to find some way of distinguishing between the two competing hypotheses. In the flocking case there are no other good competitors for the simple rules hypothesis and the hypothesis is consistent with what we know about birds. From these examples we see that we can examine the results of simulations in the same way as we examine the success of explanations.

To this point, we have examined the logical structure of a number of thought experiments. We have seen that thought experiments may employ idealized real situations or analogous fantastic situations. We have seen that thought experiments can be expressed as arguments and that they use deductive or inductive reasoning. Some thought experiments use a combination of processes.

Computer simulations test the explanatory potential of a theory or suggest causal links between one state of affairs and another. They also employ argument by analogy. We have seen that the arguments by analogy must be carefully examined for potential

disanalogies and confounding variables. The explanations that are tested by thought experiments and simulation experiments must be grounded. A grounded explanation is an explanation that is in a particular relationship to its theoretical context. In §5.4, I examine the grounding of explanations given by thought experiments and simulation experiments in detail.

2.6 Non-Thought Experiments

Before proceeding further, it is instructive to examine some examples of non-thought experiments. We have considered the utilitarian thought experiment about the young man who is asked to donate his organs to five needy people. Here is a description of a similar situation, however, it is not clear whether it is a thought experiment or not. In *The Survival Lottery*, Harris writes:

Y and Z put forward the following scheme: they propose that everyone be given a sort of lottery number. Whenever doctors have two or more dying patients who could be saved by transplants, and no suitable organs have come to hand through "natural" deaths, they can ask a central computer to supply a suitable donor. The computer will then pick the number of a suitable donor at random and he will be killed so that the lives of two or more others may be saved. No doubt if the scheme were ever to be implemented a suitable euphemism for "killed" would be employed. Perhaps we would begin to talk about citizens being called upon to "give life" to others. With the refinement of transplant procedures such a scheme could offer the chance of saving large numbers of lives that are now lost. Indeed, even taking into account the losses of the lives of the donors, the numbers of untimely deaths each year might be dramatically reduced, so much so that everyone's chance of living to a ripe old age might be increased. If this were to be the consequence of the adoption of such a scheme, and it might well be, it could not be dismissed lightly.¹⁰⁶

This example shows the importance of the context of a thought experiment. Just as we cannot always decide if someone is arguing for a point or simply describing something without examining the context of their expression, here we do not know whether the imaginary situation described in *The Survival Lottery* is a thought experiment

¹⁰⁶ Harris [1975], 89-90.

until we see how it fits into the context of the paper as a whole. In fact, there are two ways to understand Harris' hypothetical situation:

1. It is not a thought experiment. It is descriptive. Harris uses an imaginary scene to demonstrate the merits of the utilitarian position. He is not testing the utilitarian principle.
2. It is a thought experiment. Harris' approach is rhetorical. He pretends to support the utilitarian principle with a situation that is likely to be unacceptable to his audience in order to lead them to question it. In this case we have a *reductio ad absurdum*.

Utilitarianism is assumed and an undesirable outcome is obtained. The audience is then led to question the assumption. The result of this questioning has repercussions for utilitarianism.

Rawls' description of bargaining in the original position, behind the veil of ignorance is another example of a thought scenario which is a thought experiment on one interpretation and not on the other.¹⁰⁷ Rawls proposes that we imagine ourselves bargaining for the kind of society of which the people we represent will be members. We are behind a veil of ignorance in that we have no knowledge of what will be the social position, natural abilities, health, and status of the members of the future society whom we represent.

There are two conflicting interpretations of the role of the original position. One is that it is a methodological device that helps observers achieve an impartial vantage point allowing them to perceive ethico-political facts distinctly.¹⁰⁸ On the other reading, which Rawls favours, "the original position is a 'device of representation.' Specifically, it represents, in the veil of ignorance, widely accepted principles for the choice of principles

¹⁰⁷ Rawls [1971], 11f.

¹⁰⁸ D'Agostino [1997], "Reflective Equilibrium".

of justice."¹⁰⁹ The veil of ignorance veils just that information which a particular community agrees is irrelevant to choosing principles of justice.

On the first reading, Rawls does not use the original position as a thought experiment. The original position provides a way for one to perceive principles of justice that may be obscured by the types of information that is available to one who is not behind the veil of ignorance. This is not a thought experiment because it does not test a question against a theoretical context. Rather, it is a kind of device enabling clear perception.

On the second reading, the original position is a thought experiment. Rawls wants to explore conditions for forming adequate conceptions of justice. He constructs an imaginary situation in which certain information is not available to the people who are bargaining for the principles of justice which will govern the members of the community whom they represent. The situation gives the conditions for providing a conception of justice in which the information that is precluded by the veil of ignorance does not play a role. The thought experiment, then, tests this conditional:

If we could formulate our decisions about concrete and specific matters of justice from behind a veil of ignorance, then these rules would have a higher probability of conforming to the community's sense of justice, both with respect to specific matters and general principles.

Depending on the response to the thought experiment, it is confirmed or disconfirmed. Many people believe that the conditions for choosing the principles of justice should reflect the community's understanding of what kind of information is irrelevant to those choices. Through Rawls' thought experiment we calibrate, or test, our pre-theoretical beliefs about justice against our more theoretical beliefs about justice.

¹⁰⁹ D'Agostino [1997], "Reflective Equilibrium".

The original position is part of the reflective equilibrium process which can be summarized this way:

- We articulate the concept of justice which is widely accepted within a given community.
- We so devise the veil of ignorance that it embodies this concept.
- We consider what implications about concrete and specific matters of justice rational calculators standing in a trustee relation would reach subject to the particular restrictions on their calculations represented by *this* veil of ignorance.
- We compare these implications with individuals' considered judgements of justice about these more concrete and specific issues.
- When there is divergence between implications and judgments, we consider whether individuals might be willing to alter their judgments to bring them into line with principles which, after all, already express their own more abstract views about the concept of justice.
- If there is residual divergence, we modify the veil of ignorance to minimize this divergence.¹¹⁰

The process results in equilibrium when there is minimal divergence between a particular individual's considered judgements of justice about specific matters and the more general conception of justice. Later, in Chapter 4, I explain how the idea of reflective equilibrium is used in contextualist ethics.

Here is something that resembles a thought experiment, but is not. Louis Pascal, in a speech on overpopulation to a group of college students, describes the following situation:

... please imagine yourself to be in an ancient country which is ruled over by an evil king who has absolute power of life or death over all his subjects including yourself. Now this king is very bored, and so for his amusement he picks 10 of his subjects, men, women, and children, at random as well as an eleventh man who is separated from the rest. Now the king gives the eleventh man a choice; he will either hang the 10 people picked at random and let the eleventh man go free, or he will hang the eleventh man and let the other 10 go free. And the eleventh man must decide which it is to be.¹¹¹

At this point Pascal asks:

¹¹⁰ D'Agostino [1997], "Reflective Equilibrium".

¹¹¹ This and the following passages are from Pascal [1980], 106f.

I will simply ask you to hold up your hand and answer yes if you are any more than 50 per cent certain that you would make that sacrifice.

He specifies that he is not asking what the eleventh man should do but what each member of the audience would do. Then, Pascal modifies the situation:

The king says he will let his people go if you will agree to give him all the money you have and all the money you will make in the future, except of course enough for you to feed and house yourself and take care of all the absolute necessities. In other words he's asking you to be poor but not so poor that it impairs your health in any way.

Pascal then asks for a show of hands of those who feel that they are 95 per cent certain they would make the sacrifice. So far, it is clear that Pascal is not conducting a thought experiment but a poll.

Next, Pascal compares his hypothetical situation with the one in which, he claims, most people in developed countries find themselves. He writes:

Thus if you decide to go on with the life you were probably planning to lead, you will be letting 10 people die rather than give up your colour television and your cocktail parties. And that is more than gluttony, that is murder.

At this point Pascal is no longer conducting a poll but rather using the results of the poll in order to lead the members of the audience through a personal procedure for assessing their actions in relation to their intuitions. It is not a thought experiment because Pascal is not testing a question relative to a theoretical context.

Maxwell's demon is often called a thought experiment. The molecule sorting demon is proffered by Maxwell as a way to understand his molecular-kinetic theory of heat. A gas is a collection of molecules which are governed by Newton's laws. The second law of thermodynamics states that entropy must remain the same or increase in any change of state. Since gases are made up of large numbers of molecules, the law is given statistically, and so there is a small probability that the entropy in the system could decrease. Maxwell used the idea of a molecule sorting demon to allow his readers to imagine such a possibility. The demon is at an aperture that serves as a connection for

two containers of gas, one hot and the other cold. When the aperture is opened the demon sorts the molecules in such a way that entropy decreases; that is, he lets the swiftest molecules from the cold chamber into the hot chamber and the slowest molecules from the hot chamber into the cold chamber. With this demon Maxwell hoped to answer the objection to the statistical form of the laws of thermodynamics which govern gases.

Maxwell gives the thought experiment this way:

One of the best established facts in thermodynamics is that it is impossible in a system enclosed in an envelope which permits neither change of volume nor passage of heat, and in which both the temperature and the pressure are everywhere the same, to produce any inequality of temperature or of pressure without the expenditure of work. This is the second law of thermodynamics, and it is undoubtedly true as long as we can deal with bodies only in mass, and have no power of perceiving or handling the separate molecules of which they are made up. But if we conceive a being whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are still as essentially finite as our own, would be able to do what is at present impossible to us. For we have seen that the molecules in a vessel full of air at uniform temperature are moving with velocities by no means uniform, though the mean velocity of any great number of them, arbitrarily selected is almost exactly uniform. Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole so as to allow only the swifter molecules to pass from A to B, and only the slower ones to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics.¹¹²

Of Maxwell's Demon, Brown writes:

The point of Maxwell's demon is not to prove a conclusion hitherto unestablished, but instead to provide us with that elusive thing, *insight* and *understanding*. After the demon thought experiment we see how something is possible.¹¹³

Maxwell's demon does not result in a new conclusion, but illustrates a theory. It has no repercussions for the theoretical background, it simply illustrates it with an imaginary scenario. I cannot call this "a thought experiment". It does not add to, subtract from, or

¹¹² Maxwell [1872], 308-309.

¹¹³ Brown [1993b], 274.

revise statements in the theoretical context. It does not test a statement through varying variables.

Brown claims that thought experiments work in many different ways, just as real experiments do.

For example, real experiments sometimes test (i.e., confirm or refute) scientific conjectures; sometimes they illustrate theories or simulate natural phenomena; and sometimes they uncover or make new phenomena. ... Thought experiments are at least as richly diverse in their uses as this.¹¹⁴

This is a tempting and plausible view. It seems evident that the imagination is capable of richly diverse thought experimentation. However, I suggest that thought experiments may be more constrained than real experiments. In real experiments one may stumble upon some phenomenon that suggests new hypotheses for testing. In thought experiments stumbling on new phenomenon is not common. I have given examples of thought experiments that test scientific, or other theoretical conjectures. On Brown's account, Maxwell's demon is a thought experiment which illustrates a theory. I do not want to call these illustrations thought experiments. In real experiments, when we feel we have illustrated a theory, we do it through simulating a natural phenomenon, or uncovering a new phenomenon. In the latter case, there is a modification in theory because we know that a physical event has been instantiated. In a thought experiment, if we say it illustrates or simulates natural phenomena, we do not modify or create a new theory, we simply understand an existing theory better.

Schrödinger's cat, like Maxwell's demon, is not a thought experiment. It does not test a question relative to a theory. It is used to illustrate a theory by magnifying the predictions of quantum theory from the microphysical world to the physical world. In Schrödinger's scenario, a cat is in a sealed box which contains a radioactive material,

¹¹⁴ Brown [1991a], 33.

which, when it emits a particle, will trigger the release of a poisonous gas. According to quantum theory an atom exists in a superposition of states until observed. Until observed, a particle is neither in an excited, nor a ground state. If this is so, then, the cat is *neither* dead nor alive until observed. However, when we look in the box, the cat is *either* dead or alive. This experiment serves to underline the counter-intuitive principle that at the quantum level, things are in an indeterminate state until an observation is made.¹¹⁵ Schrödinger illustrates a principle of quantum theory by using things more familiar to us than particles. However, illustrating a theory is not experimenting with respect to that theory.

In the previous section (§2.5.5) I drew a distinction between simulation experiments and simulation demonstrations. Simulation demonstrations such as flight simulators are not simulation experiments for the same reason that Maxwell's demon and Schrödinger's cat are not thought experiments. These simulations demonstrate a state of affairs based on a theory. They are tools for learning or arriving at implications of a theory. They do not test the explanatory potential of the theory.

From these examples of non-thought experiments, and non-simulation experiments, we can see that not all hypothetical situations are thought experiments or simulation experiments. In order to qualify as a thought experiment or a simulation experiment a hypothetical situation must display the features of experiments I have outlined earlier in this chapter.

¹¹⁵ This experiment is described in Brown [1991a], 24-25.

3. Thought Experiments: The Grounding Problem

3.1 Grounding

Prior to this point we have focussed on the "experiment" part of thought experiments. It is now time to turn our attention to the "thought" part. The main concern of this dissertation is how an experiment which occurs in thought can lead to new and accurate conclusions about the world outside thought. This is the grounding problem.

In thought experiments, we use imagination to explore a question within a given theoretical context. The grounding conditions outlined in this chapter help us to keep the thought experiment properly related to this context. As much as the theoretical context, in turn, is grounded in the data representing the world and that data is grounded somehow in the world, the thought experiment will be grounded.

In the following well-known passage from one of the first papers on scientific thought experiments, Kuhn considers an aspect of the grounding problem:

[In] a real thought experiment, the empirical data upon which it rests must have been both well-known and generally accepted before the experiment was even conceived. How, then, relying exclusively upon familiar data, can a thought experiment lead to new knowledge or to a new understanding of nature?¹¹⁶

Kuhn answers this question with the observation that since our concepts reveal something about the world itself and about our conceptual apparatus, a thought experiment teaches us both about the world and about our concepts.¹¹⁷ In order to achieve this result, Kuhn stipulates that in thought experiments one should employ concepts as one has previously. He writes:

¹¹⁶ Kuhn [1964], 241.

¹¹⁷ Kuhn [1964], 253.

The imagined situation must be one to which the scientist can apply his concepts in the way he has normally employed them before.¹¹⁸

Here, I expand on Kuhn's idea. However, I do not claim that all the concepts must be used in their normal way. We will see that we decide which concepts must meet this condition on the basis of the role that the concept plays in the thought experiment.

Kuhn is interested in the inconsistencies which are highlighted in thought experiments. Sorensen specifies that these inconsistencies arise from vague concepts that work in most instances but are in conflict in some.¹¹⁹ Through examining the conflict highlighted by the thought experiment, the concept can be modified so that it becomes more precise. In this way we clarify our concepts and learn something new about the world. Kuhn's example of this process comes from examining children's notion of "faster" as it is tested by Piaget.¹²⁰ Children of 6 to 8 years were found to determine which of two cars was "faster" by noting which car reached the goal first and which car looked the most blurry. Piaget created a situation in which a more blurry car reached the goal last. After some puzzlement, some children were able to modify their beliefs about "faster" to reflect the difference between reaching the goal first and the velocity of the car. In this situation the children both modified their conception of speed and learned a fact about the movement of things in the world; that is, that the faster, more blurry, car does not always finish first. Out of considering the inconsistency, a vague concept is modified. Sorensen notes that it is not the vague concept itself which is inconsistent, it is the concept as applied to a certain state of affairs that yields the inconsistency.¹²¹

¹¹⁸ Kuhn [1964], 242.

¹¹⁹ Sorensen [1992a] and [1991] examine conflict vagueness in depth, and consider various strategies used to respond to it.

¹²⁰ Kuhn [1964], 242-246.

¹²¹ Sorensen [1992a], 170.

Kuhn's account of the changing of conceptual apparatus through thought experiments is similar to his account of scientific revolutions. In thought experiments new data that sheds light on an old concept causes a reconceptualization which is more consistent with the new data. In a scientific revolution, a crisis in the old scientific paradigm eventually brings forth a new one. Scientific anomalies are at first suppressed or somehow made to fit within the old theoretical framework. Then

that crisis ends only when some particularly imaginative individual, or a group of them, weaves a new fabric of laws, theories, and concepts, one which can assimilate the previously incongruous experience and most or all of the previous assimilated experience as well.¹²²

Kuhn compares this situation and that of thought experiments: "A crisis induced by the failure of expectation and followed by a revolution is at the heart of the thought-experimental situations we have been examining."¹²³

Kuhn puts two constraints on thought experiments. The first is that the thought experiment "must allow those who perform or study it to employ concepts in the same ways they have been employed before."¹²⁴ The second is that it must

present a normal situation, that is, a situation which the man who analyzes the experiment feels well equipped by prior experience to handle. Therefore, if the experiment depends, as it must, upon prior experience of nature, that experience must have been generally familiar before the experiment was undertaken.¹²⁵

Kuhn notes that the thought experiment need not be subject to a condition of overall verisimilitude. As long as normal cues are present, and the situation is relevant, the thought experiment can reveal something about our concepts and the world.¹²⁶ For example, in the car experiment with the children, if the children were asked if a tree or

¹²² Kuhn [1964], 263.

¹²³ Kuhn [1964], 263.

¹²⁴ Kuhn [1964], 252.

¹²⁵ Kuhn [1964], 252.

¹²⁶ Kuhn [1964], 246.

cabbage were faster, they would not learn about their use of concepts because the examples wouldn't give them normal cues or be relevant to their concept of "faster".¹²⁷ However, if the movements were presented in a cartoon, but still faithful to the way things move in nature, the thought experiment would succeed.¹²⁸ Like Kuhn, I hold that verisimilitude is not an overarching condition. There are certain parts of the thought experiment which are subject to the condition of verisimilitude and there are other parts which are exempt. Later in this chapter, I apply certain conditions to certain types of propositions that make up the thought experiments. Those that relate to the dependent variable must follow conditions similar to those that Kuhn lists.

Our understanding of concepts comes from the understanding of sentences which contain them. In Quine's terminology, "Observation sentences are the link between language, scientific or not, and the real world that language is about."¹²⁹ This is what we are looking for, a link between the sentences of the thought experiment and the world. However, the sentences of the thought experiment are not observation sentences. An observation sentence is defined as "an occasion sentence on which speakers of the language can agree outright on witnessing the occasion."¹³⁰ Thought experiments do not directly report observations. Observation sentences and theoretical sentences are linked by the vocabulary that they share. So Quine writes:

It is precisely this sharing of words, by observation sentences, that provides logical connections between the two kinds of sentences and makes observations relevant to scientific theory. Retrospectively those once innocent observation sentences are theory-laden indeed. An observation sentence containing no word more technical than 'water' will join forces with theoretical sentences containing terms as technical as 'H₂O'. Seen holophrastically, as conditioned to stimulatory situations, the

¹²⁷ Kuhn [1964], 245.

¹²⁸ Kuhn [1964], 246.

¹²⁹ Quine [1990], 5.

¹³⁰ Quine [1990], 3.

sentence is theory-free; seen analytically, word by word, it is theory-laden.¹³¹

Our understanding of concepts, then, arises from our understanding of sentences.

Observations belong to the context of discovery; concepts to the context of justification.

On Kuhn's account the thought experiment can bring out inconsistencies between the concepts and the network of sentences within which they occur. In the following passage, Kuhn describes how an historian studies the use of concepts:

[S]cientific concepts are invariably encountered within a matrix of law, theory and expectation from which they cannot be altogether extricated for the sake of definition. To discover what they mean the historian must examine both what is said about them, and also the way in which they are used. In the process he regularly discovers a number of different criteria which govern their use and whose coexistence can be understood only by reference to many of the other scientific (and sometimes extrascientific) beliefs which guide the men who use them. It follows that those concepts were not intended for application to any possible world, but only to the world as the scientist saw it. Their use is one index of his commitment to a larger body of law and theory. Conversely, the legislative content of that larger body of belief is in part carried by the concepts themselves.¹³²

The idea that the concepts should be used in their normal way is related to grounding. Concepts are formed with respect to the data, conventions and theoretical contexts. Thought experiments teach us something about the data, conventions and theoretical context when concepts (except those that are deliberately idealized or simplified, or those that are varied in the independent variable) are used in their normal way.

In my analysis of the normal use of concepts I examine how the concepts used in thought experiments must conform to their use in the sentences in which they normally occur. I also examine the propositions that make up the thought experiment.¹³³ A

¹³¹ Quine [1990], 7.

¹³² Kuhn [1964], 259-260.

¹³³ In this chapter, I refer to "sentences", "propositions" and "statements". I use these terms in the following way: A sentence is any grammatically structured group of words. A proposition is a sentence which can be assigned a truth value. A statement is a verbally asserted proposition.

thought experiment may reveal an inconsistency in the way that the concept is applied, and it may also reveal an inconsistency between various propositions that describe that thought experiment and its relationship to its theoretical context, which includes the primary theory and background theories and assumptions.

Now, consider the following passage given by Wittgenstein:

Imagine, for example, that all human bodies which exist looked alike, that on the other hand, different sets of characteristics seemed, as it were, to change their habitation among these bodies. Such a set of characteristics might be, say, mildness, together with a high pitched voice, and slow movements, or a choleric temperament, a deep voice, and jerky movements, and such like. Under such circumstances, although it would be possible to give the bodies names, we should perhaps be as little inclined to do so as we are to give names to the chairs of our dining-room set. On the other hand, it might be useful to give names to the sets of characteristics, and the use of these names would now *roughly* correspond to the personal names in our present language.¹³⁴

Here, Wittgenstein makes careful reference to the way that our use of language is determined by its context. If things were very much different from how they in fact are, we would adjust our language. Kuhn's insistence on ordinary usage must come from concerns like these. In fact, most thought experiments have a tacit *ceteris paribus* clause which has the function of holding the context of the concepts steady. This passage illustrates the strong connection between terms and their theoretical context. This connection is vital to grounding.

Kuhn is interested in how thought experiments can lead to new knowledge. I am interested in how we ensure that the new knowledge is connected to the world in an accurate way; that it is, in fact, new knowledge. Kuhn claims that concepts are legislated by the theoretical context and the concepts themselves. It is to theoretical context and conceptual analysis that I now turn in order to explicate my account of grounding.

¹³⁴ Wittgenstein [1958a], 61-62.

3.2 Grounding Propositions

In order to examine grounding, it is useful to consider various accounts of the grounding of propositions. Consider the Epimenides sentence:

All sentences asserted by Cretans are false.

Is this sentence true, or false, when asserted by Epimenides, a Cretan? If we think that this sentence is true, then it must be false, since it is asserted by a Cretan. If we think that the sentence is false, then it must be true, since it is asserted by a Cretan. If we accept classical bivalence, we cannot have a sentence that is both true and false, or neither true nor false. A paradox ensues.

Various strategies have been suggested to solve paradoxes such as that associated with the Epimenides sentence. Kripke develops Herzberger's idea that a sentence which states that another sentence is true will be true or false in relation to the truth value of the sentence to which it refers. If this next sentence states that a third sentence is true, it will, in turn, be true or false depending on the truth value of that sentence. The original sentence will be grounded if this process ends with a sentence which does not use a truth predicate.¹³⁵ That is, the final sentence, the sentence in which the series of sentences referring to other sentences ends, must not assert the truth or falsity of some other thing or sentence, but rather must state a fact about the world. Herzberger writes that "there is something schematic in the concept of truth, which requires filling in."¹³⁶ Herzberger compares the search for a ground for a paradoxical sentence to be like that of "the bureaucratic regress in which each clerk endlessly refers you to the next to settle your accounts."¹³⁷ If a sentence invokes another sentence and that sentence yet another, there must be a sentence at the base level of the chain which does not contain a truth value, but

¹³⁵ Irvine [1992], 283.

¹³⁶ Herzberger [1970], 150.

¹³⁷ Herzberger [1970], 150.

is true or false in relation to the world. Herzberger claims that sentences that are not grounded in this way may have a sense, but they have no truth value and so determine no statement.¹³⁸

Kripke illustrates the role of empirical facts in deciding whether statements involving the notion of truth are paradoxical or not. He asks us to consider the following statements.¹³⁹ The first is made by Jones:

(1) Most (i.e., a majority) of Nixon's assertions about Watergate are false.

The second is made by Nixon:

(2) Everything Jones says about Watergate is true.

The truth value of the first statement can be determined by examining the truth or falsity of all of Nixon's statements about Watergate. However, if we find that Nixon's statements about Watergate, excepting (2), are half false and half true; and Jones has only made statement (1) about Watergate, then a paradox obtains.

On Kripke's account, we see that paradoxical sentences such as the ones above cannot be assigned truth values at any point. As a result, Kripke holds that there are truth-value gaps; that is, contrary to the idea of classical bivalence, there are some sentences that are neither true nor false.¹⁴⁰

Kripke uses the Watergate example to underscore his observation that

an adequate theory must allow our statements involving the notion of truth to be *risky*: they risk being paradoxical if the empirical facts are extremely (and unexpectedly) unfavorable. There can be no syntactic or semantic "sieve" that will winnow out the "bad" cases while preserving the "good" ones.¹⁴¹

¹³⁸ Herzberger [1970], 152.

¹³⁹ Kripke [1975], 691.

¹⁴⁰ Kripke [1975], 698.

¹⁴¹ Kripke [1975], 692.

Irvine's concern is to link language and belief more firmly to the empirical world. Irvine holds that since language is constrained from only two directions — social/linguistic convention and the world itself — and since nature contains no contradictions, contradictory beliefs must be the result of inappropriate social/linguistic conventions. Paradoxes like the one above arise when we assume, through such conventions, that sentences which are ungrounded are nevertheless meaningful. When it becomes clear that these types of sentences are ungrounded, Irvine concludes that they are merely non-linguistic things in the world, unable to communicate meaning.¹⁴²

The sentences of a thought experiment will be, or will fail to be, grounded in a theoretical or empirical context. We can only go so far in stipulating the form of a successful thought experiment and then we need to check with its theoretical context. We need to ask if there is a theory or empirical state of affairs to which the sentences of the thought experiment refer (Herzberger). We also need to ask how this theory or state of affairs relates to the sentences of the thought experiment (Kripke). Finally, we can ask whether there are distortions arising from social/linguistic conventions that upset the grounding link (Irvine). The first two issues are addressed in the remainder of this chapter. The last, I examine in Chapters 6 and 7.

3.3 Explicating Grounding

Consider my claim that a thought experiments has two parts: the experimental set-up and the responses to the experimental set-up. Both the experimental set-up and the responses must be grounded in the theoretical context of the thought experiment. In this chapter, I deal with grounding the experimental set-up. In Chapters 6 and 7, I deal with grounding responses to thought experiments. Let us begin by considering two components of the experimental set-up, concepts and theoretical contexts.

¹⁴² Irvine [1992], 298.

3.3.1 Concepts

In a grounded thought experiment concepts must be defined normally. This relates to Kuhn's stipulation that concepts used in a thought experiment must be employed in a normal way.

In *Human Beings*¹⁴³ Johnston observes that unspecified concepts make many thought experiments about personal identity suspect. He writes:

We have found good reason to suspect intuitions that show that one or another form of continuity is not necessary for survival; for such intuitions may simply be generated by an unspecific concept of ourselves as some or other kind of locus of mental life.¹⁴⁴

As an alternative to this unspecific and problematic conception of ourselves, Johnston suggests that we are better to have a naturalistic conception of ourselves as human organisms who have non-problematic ways of recognizing and identifying one another on the basis of our continuous mental and physical functioning.¹⁴⁵

Johnston claims that the method of theorizing about personal identity by appeal to our intuitive reactions to thought experiments would be justified if the following two requirements were satisfied:

First, ... the concept of being the same person should be able to be represented as necessary and sufficient conditions for application of the predicate "is the same person", conditions that could be cast in terms of statements about continuity and dependence, statements not themselves to be explained in terms of statements about personal identity. ... Secondly, our intuitive reactions to the puzzle cases should be able to be taken as manifestations of our grasp of those necessary and sufficient conditions, and not as overgeneralizations from the everyday run of cases or manifestations of a particular concept of people ...¹⁴⁶

¹⁴³ Johnston [1987].

¹⁴⁴ Johnston [1987], 80.

¹⁴⁵ Johnston [1987], 75.

¹⁴⁶ Johnston [1987], 60.

Johnston claims that there is a dominant philosophical view about personal identity which is defended by the method of thought experiments. Rather than viewing intuitive reactions described in terms of continuity and connectedness as a manifestation of our concept of personal identity, he suggests we regard these things as evidence for identity and work on specifying our concepts in a more precise and non-question-begging way.¹⁴⁷

I agree that we need to examine whether the concept of a person with which we enter into our thought experiments is too unspecific, or too unexamined, to lead us to interesting responses. However, if we can represent the concept of being the same person with a complete set of necessary and sufficient conditions, there would be little reason to conduct a thought experiment to explore the concept further.

When I claim that concepts used in thought experiments should be defined normally, I do not mean that they should be analytically defined, nor do I mean that we should have a complete set of necessary and sufficient conditions to define them. Rather, the notion of a concept that is defined normally is a pragmatic notion. We use thought experiments to clarify our usage of concepts. Concepts are understood within sentences; we must explore the concept as it is normally used in the sentences which make up the theoretical context of the thought experiment. The notion, "defined normally", is a pragmatic notion linking our thought experiments with our use of language and our operations in the world.

So, there is something right about Johnson's suggestion that we look to our non-problematic practice of recognizing and identifying one another over time as a starting point for our enquiry. Ideally, a personal identity thought experiment will explore the limits and nuances of our current conceptions as they are embodied in paradigm cases

¹⁴⁷ Johnston [1987], 71.

such as those Johnston suggests. From this unremarkable but plausible starting point, implicit beliefs, inconsistencies, contradictions and new conceptual possibilities may be discovered.

Applying concepts in the manner in which they are normally used is a condition of grounded thought experiments. On the other hand, concepts cannot be useful, even if they are used in their normal way, if they are used in a shifting context. That is, in a shifting context, there is no normal way to use concepts. In the following passage Wittgenstein describes the problem of applying concepts that are not clearly prescribed by facts:

It is only in normal cases that the use of a word is clearly prescribed; we know, are in no doubt, what to say in this or that case. The more abnormal the case, the more doubtful it becomes what we are to say. And if things were quite different from what they actually are — if there were for instance no characteristic expression of pain, of fear, of joy; if rule became exception and exception rule; or if both became phenomena of roughly equal frequency — this would make our normal language-games lose their point. The procedure of putting a lump of cheese on a balance and fixing the price by the turn of the scale would lose its point if it frequently happened for such lumps to suddenly grow or shrink for no obvious reason.¹⁴⁸

Here we see that concepts expressed by words are understood in relation to the way the world is. Wittgenstein claims that if the facts were to change, the words would not be useful, since it would be hard to know how to apply them.

Nowell-Smith has a similar point to make in discussing desert-island thought experiments. The thought experiment he discusses is one in which there are two inhabitants on a desert island. As A dies, he elicits from B a promise to follow certain instructions in disposing of A's goods should B reach home. Is B under an obligation to keep the promise, if B finds his way home and decides that there is a more beneficial way to dispose of the goods? Remember that no-one but B is aware of the promise.¹⁴⁹ There

¹⁴⁸ Wittgenstein [1958b], 142.

¹⁴⁹ Nowell-Smith [1954], 240.

seem to be only good effects, and no ill-effects, incurred by breaking the promise.

Nowell-Smith writes:

In the same way I confess to being quite unable to decide *now* what I should say if a desert-island situation arose. Moral language is used against a background in which it is almost always true that a breach of trust will, either directly or in the more roundabout ways which utilitarians suggest, do more harm than good; and if this background is expressly removed my ordinary moral language breaks down.¹⁵⁰

Our concepts are understood through our understanding of clusters of sentences describing the way things normally are. In order for a concept used in the thought experiment to be employed in its normal way, it must conform to its use in the propositions of the theoretical context. In order to understand this idea more clearly, consider two types of concepts, 1, empirical concepts and 2, value concepts.

1. Empirical concepts: Wilkes writes that science is concerned with natural kind concepts, in that it picks out groups of things that are usefully isolated. She defines natural kind terms as "the terms for which, and with which, the laws and generalizations of science are framed."¹⁵¹ Philosophical difficulties with a strong notion of natural kinds do not affect the claim that it is through some method of isolating groups of things under laws and generalizations that we gain the framework for determining the grounding of concepts of the thought experiment. It is the weaker notion of natural kinds as useful scientific and conceptual groupings that I am using here.

Wilkes writes that natural kind terms are "law-cluster concepts" which contain "core facts."¹⁵² The majority of the laws that define the kind, and the laws that define the core facts, need to hold if something is to qualify as a natural kind. This clarifies the process of determining what is theoretically grounded, for it is against the laws and facts

¹⁵⁰ Nowell-Smith [1954], 241.

¹⁵¹ Wilkes [1988], 14.

¹⁵² Wilkes [1988], 28.

that govern the natural kind concepts used in a thought experiment that the concepts must be checked to see whether they conform to the experiment's theoretical context.¹⁵³

2. Value concepts: Value concepts appealed to in the thought experiment must conform to their theoretical context. Value concepts are based on sets of propositions which reflect norms and values, which include: values and norms which have accrued from the history of human social experience, often considered to be common sense values and norms, and values reflecting our self-conceptions.

Our norms and values are not necessarily consistent. It is through conventions, both social and philosophical, that we build norms and values into frameworks which serve as backgrounds for our thought experiments. I explain the idea of frameworks of values in more detail in Chapter 4.

I have divided the data into two categories which match generally the fact/value distinction. I am not going to take up questions about this distinction here. I simply want to consider the least processed data from which the concepts used in thought experiments are constructed. Consider that the concepts used in thought experiments must be grounded in their data. For the physical world, the concepts must be grounded in empirical observations; for the world of values, in beliefs arising from a framework of values. Just as I do not take up the fact/value distinction, so I do not take up the realist/antirealist question, in the realm of values or facts. Empirical observations are the closest connection we have to understanding the physical world, whether we believe that they actually reflect it, give a consistent way of working with it, or something else. With respect to values, they may exist or not, the key point is the general acceptance of the considered values in the social milieu and framework of values of the thought experimenter.

¹⁵³ Wilkes [1988], 28-29.

The development of data into concepts is mediated by various background influences, such as background theories and various conventions. By background theories, I mean those theories which operate at the level of assumption. These theories affect the concepts and yet are not the main theory within which the concepts will be employed. The main theory of the thought experiment is its primary theory. For example, the primary theory of a personal identity thought experiment is that of personal identity. Our personal identity theories are, in turn, closely linked with background theories of biology and psychology. At a deeper level, the concept of personal identity is also based on other background theories including logic, physics and a number of other "commonsense" assumptions.

Conceptual development is also mediated by conventions, typically linguistic, scientific, social and psychological. These operate with various strengths in various theoretical contexts. For example, the concept of personal identity is largely mediated by social conventions. We can see this when we notice that statements like "Edward's parent is the same person who used to be Clarissa's teacher" link personal identity and social roles. The following statement links personal identity with biology and behaviour. "Since the onset of Alzheimer's, she seems to be a different person." Ethical concepts, like personal identity concepts, are similarly heavily mediated by social conventions.

Concepts are used in propositions which reflect the data from empirical observations and beliefs about values. We understand concepts by comparing the way that the concept is used in various sentences. In thought experiments, concepts are defined normally when they are used in a way that conforms to their use in the primary theory, the background theories and the related conventions of the thought experiment. The idea of concepts that are defined normally is meant to invoke the pragmatic understanding and use of concepts in their normal context. Concepts that are defined normally are those which are described by propositions that:

1. are consistent with the theories and conventions

and

2. are plausible and testable in relation to the theories and conventions,

For example, the concept of oranges is used in a cluster of propositions including:

A. Oranges are orange-ish in colour.

B. Oranges are a fruit.

C. Oranges are round-ish.

Now consider adding two other propositions to the group of propositions that use the concept:

(i) Oranges are a vegetable.

(ii) Oranges grow on many other planets in the universe.

Proposition (i) is inconsistent with proposition B, relative to most background theories which will distinguish vegetables and fruits. So proposition (i) does not conform to the concept on criterion 1.

Proposition (ii) does not seem plausible, given background theories about other planets and their climates. The probability of other planets having the type of climate that would support the growth of oranges, and oranges evolving in, or being transplanted to, that planet seem slim. Proposition (ii) does not seem testable since it is not currently possible to predict how many planets there are in the universe and, even if we could do that, it is not plausible that we would have access to them. Proposition (ii) does not conform to the concept of oranges on criterion 2.

A grounded thought experiment will use a concept that is defined normally, except when the concept is playing a certain role in the thought experiment. (This

situation is described in the following paragraph.) A concept that is defined normally is used when the group of propositions which describe it conform to the theories and conventions which constitute its normal theoretical context. If propositions (i) and (ii) were added to our concept of oranges, it would not be a concept that is defined normally in that it would not conform to the theoretical context. That is, the propositions that use the concept would not be consistent, plausible and testable in relation to the theoretical context.

There is one case in which concepts need not be defined normally. This is the case in which an aspect of the concept itself is under scrutiny. For example, in §5.1, I examine a thought experiment which explores the relationship between narrow psychological states, extensions and the meaning of the term "water". In this thought experiment, aspects relating to the meaning of the word "water" are modified in various ways in order to examine how they relate. In this case, the term "water" is not defined normally.

3.3.2 Theoretical Contexts

Grounding happens in layers. We must ground the thought experiment in the theoretical context. The propositions of the theoretical context itself must be grounded in empirical observations and beliefs about values, which are expressed in sentences using concepts. The concepts must be defined normally according to the criteria outlined above. In this way, it is hoped that they will reflect, as accurately as possible, relevant empirical data and values.

We shall see that when it is difficult to evaluate the relationship between the theoretical context and the data, it is difficult to clarify the relationship between the thought experiment and its theoretical context. That is, when the theoretical context is unclear, it is difficult to evaluate whether a thought experiment is grounded or not.

On the other hand, thought experiments are designed, in part, to clarify the very theoretical context within which they must be grounded. This clarification is achieved by designing a thought experiment to examine well-formed questions whose answers will have repercussions for its context. These questions are explored by tracking dependent and independent variables that are relevantly related to the questions and their theoretical contexts. Sometimes a thought experiment will give a result that cannot be incorporated within the existing theory. In this case, a new kind of scientific, social, or philosophical convention may be needed to interpret the new results and reorganize the data into a new theory. We have seen that Kuhn calls this phenomenon a paradigm shift when it occurs in science. One of his standard examples is the Copernican Revolution. It is called a revolution because the revelation that it is the sun, and not the earth, that is the centre of the solar system had far-reaching effects. The earlier idea that the earth is the centre of the solar system led to complex formulations of the trajectories of the orbits of the sun and the planets. As more data was collected it became more and more difficult to reconcile the data with the existing theory. The new idea that the sun is the centre simplified the theories explaining the behaviour of the planets. Moreover, it challenged and changed social and religious conventions which pertained to our place in the physical world.

Explicit propositions contained in thought experiments (except those describing extraneous particulars) need to be grounded in their theoretical context. They are grounded whenever they are relevantly related to that context. The conditions that govern how propositions of the thought experiment must be relevantly related to its theoretical context vary depending of the function the propositions play. Thought experiments are made up of propositions that describe:

1. the dependent variables
2. the independent variables

3. the simplifying assumptions

and,

4. those experimental conditions which are not simplifying assumptions and do not affect the variables.

Each type of proposition has its own grounding conditions, as follows:

1. Propositions describing the dependent variable must conform to the theoretical context of the thought experiment. By "conform" I mean that they are consistent with, relevant to, and plausible with respect to the propositions of the theoretical context. These conditions are similar to the conditions that concepts that are defined normally exhibit in relation to their theoretical contexts.

In order to consider the conforming conditions in more detail, consider two similar thought experiments that can be characterized by the following conditional sentences expressing the relationship between their dependent and independent variables:

If we could split like amoebae, then we would all be "one".

and

If we could split like amoebae, then personal identity could not be traced in terms of bodily continuity.

Recall that the dependent variable is the consequent of the conditional sentence expressing the relationship between the dependent and independent variables. In this case, the first conditional sentence describes a thought experiment that is not grounded, while the second conditional sentence describes one that is. In the first, the dependent variable, (all being "one") is not consistent with the theoretical context (personal identity theory) in that it is impossible to conceive of the idea of identity as an individuating feature if we are all somehow one. It is irrelevant to the theoretical context for the same

reason. At the same time, we might say that the theoretical context would be irrelevant to it, if we were all one. Finally, it is difficult to ascertain whether the dependent variable describes a plausible concept. Clearly, this first thought experiment is not grounded because the dependent variable does not conform to the theoretical context. That is, it is not consistent with, relevant to, or plausible with respect to the theoretical context. Another way of seeing that the dependent variable is not grounded is to note that it relates to a question which is not relevantly related to the theoretical context. It is the question under examination which helps determine the dependent variable. In this case, the question would have something to do with being "one", while the theoretical context addresses the problem of tracing the identity of individuals.

The second conditional contains a dependent variable that is grounded. The idea that there might be difficulty tracing personal identity in terms of bodily continuity is not inconsistent with current ideas of personal identity. There is an ongoing debate between those that hold some kind of physical view of personal identity and those that hold some kind of psychological view. The dependent variable, then, also expresses a proposition which is relevant to the theoretical context. The proposition is also plausible given the theoretical context. We are acquainted with people having the same identity across many stages of life and through many changes and modifications of physical form. The dependent variable is plausible given our experience and the current debates about personal identity. We say that this dependent variable is grounded since it is consistent with, relevant to and plausible with respect to, the theoretical context. of the thought experiment. That is, it conforms to the theoretical context.

2. Propositions describing the independent variable do not have to conform to the propositions that make up the theoretical context. The independent variable is varied in a way that is relevantly related to the question under scrutiny. For example, in the thought

experiment about the surgeon and his five patients¹⁵⁴ the relationship between the dependent and the independent variable is expressed in the following sentence:

If a surgeon requires the use of your body to save five others, you are morally obliged to acquiesce.

Recall that the independent variable is the antecedent of the conditional sentence expressing the relationship between the dependent and independent variables. The independent variable describes an implausible situation. We do not have surgeons making these kinds of requests. It does not conform with our background beliefs about surgeons. However, the independent variable is set to determine the applicability of the utilitarian principle. We can see that the independent variable is set in a way that is relevant to the question under examination.

The independent variable may be not only implausible, but also impossible. Consider what Sorensen writes about a thought experiment that begins: "Imagine the universe doubled in size last night." He writes:

... when philosophers suppose that the universe doubled in size last night, they tend to think that the assumption is accommodated by merely multiplying all the magnitudes by two. However, the world is governed by geometric relations such as the inverse square law for gravity. Doubling the size of objects would not preserve their surface to mass ratios. So the only way that the hypothesis could be really unverifiable is that the doubling would immediately cause all observers to black-out and die.¹⁵⁵

Wilkes has similar reservations about the amœba personal identity thought experiment:

Consider next one of the familiar thought-experiments to do with personal identity: that we might all split like amœbae. It is obviously and essentially relevant *to the purposes of this thought-experiment* to know such things as: how often? Is it predictable? Or sometimes predictable and sometimes not, like dying? Can it be induced, or prevented? Just as obviously, the background society, against which we set the phenomenon, is now mysterious. Does it have such institutions as marriage? How

¹⁵⁴ See §2.1.

¹⁵⁵ Sorensen [forthcoming].

would that work? Or universities? It would be difficult, to say the least, if universities doubled in size every few days, or weeks, or years. Are pregnant women debarred from splitting? The *entire* background here is incomprehensible. When we ask what *we* would say if this happened, who, now, are "we"?¹⁵⁶

In the second part of this quotation, Wilkes airs observations similar to those of Sorensen. In many thought experiments, the results of the supposition given as the independent variable have effects on the background conditions that, if considered, would make the relevance of the supposition to the question under consideration impossible to ascertain.

Even worse, some thought experiments appeal to conditions that are impossible to verify. For example, if I said, "Consider that your soul migrates to that of an eagle and then into a Douglas fir" and then asked about personal identity in relation to that, you would be entitled to respond that the thought experiment is meaningless. There is no known way to detect souls, soul migrations and personal identity transfers that might result from these phenomena.

However, there is a response to this problem of bizarre counterfactuals. We fill in background conditions where necessary, which is where they directly affect the dependent variables. Then, we decide to let the other, deeper assumptions go. Sorensen gives examples of cases in which the relationship of the thought experiment to its background does not matter. He writes:

... many bizarre counterfactuals are knowable by virtue of instantiating logical and semantic relationships between the antecedent and the consequent:

- (1) If everything doubled in size last night, then my pencil would not be longer than my desk.
- (2) If solipsism were true, then there would not be more than one person who understands English.

¹⁵⁶ Wilkes [1988], 11.

(3) If there were a proof that had infinitely many steps, then a finite being would not be able to survey it.¹⁵⁷

In these types of thought experiment the relationship of the thought experiment to other parts of the background does not matter. The hypothetical is designed to tease out the logical and semantic relations between things.

Another defense of counterfactuals given by Sorensen is that while bizarre counter-factuals may have repercussions if they were carried out in detail, most thought experimenters mean to present a counterfactual that applies only in the manner stipulated, while in other ways the world of the thought experiment is like the actual world.

Sorensen writes:

We wisely trade away realism in favor of radically simplified scenarios to ease calculation and to underscore relevant variables.¹⁵⁸

This is related to my treatment of simplifying assumptions which follows. We need to be clear whether, or not, the simplifying assumption affects the interaction between the independent and dependent variables. If it does, it must be represented in the conclusion.

Independent variables, then, need not conform to the propositions of the theoretical context. However, the independent variables must be varied in a way that is relevantly related to the question under scrutiny.

3. Simplifying assumptions and idealizing assumptions need not conform to the theoretical context, although they must be carried forward to the conclusion. Typically, simplifying assumptions are employed to help isolate variables. Ideally, these variables and simplifying assumptions are carefully set in relation to the theoretical context. Simplifying assumptions need not conform to the theoretical context, but their relationship to it must be set by their being associated with exploring the parameters of

¹⁵⁷ Sorensen [forthcoming].

¹⁵⁸ Sorensen [forthcoming].

the question derived from the theoretical context. Recall the following examples of simplifying assumptions:

A. In Thomson's thought experiment about the ill violinist (§2.5.1), we are asked to assume that the foetus is a person from the moment of conception. It is difficult to ascertain whether this assumption conforms with the theoretical context or not, since it varies depending on the framework of values of the individual conducting the thought experiment. However, since Thomson's assumption is a simplifying assumption, it need not conform to the theoretical context. A simplifying assumption need not conform because it is carried forward to the conclusion of the thought experiment. Through her simplifying assumption, Thomson is able to apply the results of the thought experiment to all foetuses, and the question of at what point the foetus has the status of a person, and how that relates to abortion questions, is avoided.

B. In Stevin's experiment (§2.4.2), we employ the simplifying assumption of a frictionless environment. This assumption does not conform to the theory of physics, the primary theory of the thought experiment. The simplifying assumption is carried through to the conclusion and so we account for the effect of friction when we apply Stevin's result.

4. Experimental conditions not affecting the variables are often employed to help the audience imagine the situation, or to facilitate the set-up of the situation. For example, in the thought experiment about the utilitarian surgeon given in §2.1, the fact that it is a surgeon, rather than an engineer, who is contemplating the utilitarian action is not critical to the question of whether it is right to take one life to save five. However, the described set-up is useful for the audience of the thought experiment. We can easily imagine the situation and the problem in the context of a surgeon's practice.

When we employ experimental conditions not affecting variables we need to be sure they don't affect the variable. That is, they must really involve extraneous particulars. It is important to note that theories are made up of inter-dependent laws.

Wilkes lists three ways this can be explicated:

- (a) Phenomena governed by one set of laws will frequently interact with phenomena governed by another set.
- (b) Any given law must itself be amenable to explanation at a deeper level, where 'deeper' usually means either 'more theoretical' or 'microstructural', or both.
- (c) The laws of two theories in different domains should be consistent with one another.¹⁵⁹

These three points highlight the improbability of an isolated exception to a law. If we ignore or fail to take into account a particular law that is relevant to a putatively extraneous particular of a thought experiment, the effect may be felt throughout the whole experiment, whether it is in the interaction of this law with other laws, or in its subsumption under more fundamental laws that describe the system.

It is important to distinguish between the various functions of the propositions that make up the thought experiment. Some will represent the dependent, and some the independent, variable. Some will spell out those details needed to embed the thought experiment in its theoretical context. Some will relate to the primary theory and some to the various background theories. Some propositions will involve relationships between the primary theory and the background theories. Determining the function of the propositions that comprise the thought experiment is not an easy task, but it is essential in order to decide whether the thought experiment is sufficiently grounded.

A view, like that of Wilkes, which suggests that a philosophical thought experiment is valid if it does not violate the laws of nature, is too strong. If the theoretical context of the thought experiment has nothing to do with the laws of nature,

¹⁵⁹ Wilkes [1988], 29-30. Note that (c) is an extension of (a).

violating the laws of nature will have no effect on the success of the thought experiment. Moreover, the independent variable may violate the laws of nature, if it is relevantly related to the question of the thought experiment, even when the laws of nature are relevant to the theoretical context.

In summary, the account of grounding has two parts. A grounded thought experiment uses concepts that are defined normally (except when some aspect of the concept itself is under scrutiny) and propositions (excepting propositions which describe extraneous particulars) which are relevantly related to the thought experiment's theoretical context and the question the thought experiment is designed to explore.

In the next two chapters, we see how the grounding conditions given here apply to thought experiments concerning ethics and other fields.

4. Grounding Thought Experiments in Ethics

The result of a thought experiment arises from the thought experiment's set-up and the responses to the thought experiment. In the last chapter, we characterized the nature of a grounded relationship between the thought experiment's set-up and the theoretical context. In Chapters 6 and 7, we will examine the nature of a grounded relationship between the responses to the thought experiment and their theoretical context. In this chapter, I examine the nature of the theoretical context of thought experiments concerning ethics and how this affects grounding.

Thought experiments in ethics pose a particular kind of obstacle to grounding. This problem is present in any area in which the theoretical context is unclear. We will see this problem again in Chapter 5 in relation to thought experiments concerning personal identity. Thought experiments in ethics are generally conducted in relation to what I call a "framework of values" as well as to a variety of relevant background theories. We shall see that a framework of values is not a well-defined theory, even though it is the primary part of the theoretical context of thought experiments concerning ethics.

Now, consider Russell's comment that:

[T]hose questions which are already capable of definite answers are placed in the sciences, while those only to which, at present, no definite answer can be given, remain to form the residue which is called philosophy.¹⁶⁰

The questions of philosophy are those which explore areas of interest which are not as yet sharply defined. In ethics, both the questions raised and their relationship to the developing theoretical context have not been clearly enough understood for the questions to become scientific ones. For example, Irvine writes:

¹⁶⁰ Russell [1912], 90.

In one sense, then, philosophy and science are on a par; both have as their goal our understanding of the world. At the same time, philosophy is pre-scientific in the sense that the questions and issues it addresses are often not yet well-defined, not sharp enough to be regularized into normal science. Philosophical analysis is what gives us this sharpening.¹⁶¹

and later,

Once we know, in principle, how to answer a question, that question is, by definition scientific. Questions which we do not, even in principle, know how to answer (either because they are ill formed or too vague, or not sufficiently connected to the world) remain, as we might expect, pre-scientific.¹⁶²

The interesting thing about the distinction between pre-science and science is that it highlights the different level of certainty regarding how to answer scientific and pre-scientific questions. This certainty is a reflection of the development of the theoretical context. In many areas of science the context is strongly developed. In many areas of philosophy, such as personal identity theory and ethics, the theoretical background is not as clear.

Arguing in a similar vein, Wilkes writes that:

In science we often have a supporting background of theory, and experiments conducted against fixed physical parameters; thus in many cases the situation is indeed adequately described, and relevance or irrelevance can be determined with fair safety. When we move away from the relative security of the mature sciences to the rich and glorious chaos of common sense — which is the domain in which most philosophical thought experiments romp around — this is missing.¹⁶³

In thought experiments in philosophy, such as those in ethics and personal identity theory, there is less definition in the ground of the thought experiment. When the theory is less well developed, it is more difficult to characterize aspects of the relationship of the experimental situation to its background.

¹⁶¹ Irvine [1997], 84.

¹⁶² Irvine [1997], 86.

¹⁶³ Wilkes [1988], 13.

Sorensen also argues that philosophical thought experiments deal with questions that have a less well-defined relationship to their subject matter than those in science. He writes:

Philosophy is essentially protoscientific. It speaks to the miscellany of questions that cannot be addressed in standardized ways. One of the philosopher's jobs is to develop approaches that have some prospect of becoming a research policy for these homeless questions. When successful in a small way, the question is pushed into another, better established field. ... the difference between philosophy and science lies in our *relationship* to the subject matter, not in the topics themselves. Just as news is constituted by the external requirement of being recently known, philosophy is under the external constraint of being methodologically unsettled. Hence, the conceptual disarray that is so symptomatic of philosophy does not confine its thought experiments to the job of testing analytic propositions.¹⁶⁴

Massey argues that the methodology of thought experimentation needs to be developed in a more rigorous manner. Rather than relying on the thesis of facile conception — that whatever can be conceived is possible — we need to develop rigorous methodological standards for thought experiments.¹⁶⁵ He notes that when meaning analysis was discarded due to Quine's dissolution of the analytic-synthetic distinction, analyticity was revived in the form of conceivability arguments, or thought experiments.¹⁶⁶ He charges analytic philosophers with not bothering to question what standards of conceivability to apply to thought experiments. By contrast, he cites examples of theories of conceivability and appropriate conceivability standards that are developed for scientific reasons such as Hermann von Helmholtz's theory of spatial conception.¹⁶⁷ Massey's paper leaves the reader with the suggestion that standards for thought experiments need to be developed that have the rigour of those governing scientific methodology.

¹⁶⁴ Sorensen [1992a], 94-95.

¹⁶⁵ Massey [1991], 288f.

¹⁶⁶ Massey [1991], 287-288.

¹⁶⁷ Massey [1991], 294.

Massey's point is well to consider, we need to develop rigorous standards for thought experiments. This is the motivation for this dissertation. However, Sorensen's quotation above suggests that philosophy is methodologically unsettled partly because it deals with questions that cannot be addressed in standardized ways. When the questions become better defined and more clearly understood they move to a more established field with more standardized methodology.

With respect to thought experiments in philosophy, and particularly in ethics, the grounding issue is more complex than my account in the previous chapter may lead one to believe. Sorensen stresses the relationship of the thought experiment to its theoretical context. The relationship is less clear and more exploratory in philosophical fields than in scientific ones. In ethics, the theoretical context is both developing and closely linked with a number of disparate background considerations. To understand how we can make progress with the grounding problem for thought experiments concerning ethics, it will help to examine, first, the contextualist approach to ethics.

4.1 Contextualism

One prominent form of contextualism views morality as a social construct arising within a particular cultural and historical context. Contextualism is practically oriented, in that it is concerned with the way that ethical problems are experienced and resolved in particular situations. On this account, a problem is considered in light of the various parameters and constraints that bear on it, including abstract moral principles, the social or institutional setting of the problem, and any background theories that are pertinent to the problem, such as psychology or medicine.¹⁶⁸ A number of possible outcomes are considered and the outcome (or one of the outcomes) that best balances the multiple constraints bearing on the situation is selected.

¹⁶⁸ Winkler [1996], 52.

How is it that multiple constraints bearing on the problem are balanced and solutions are proposed? The answer comes from the idea of wide reflective equilibrium. This idea derives from Rawls' work, in which he suggests a process of reflective equilibrium in order to bargain for the principles of justice that would apply to a particular society.¹⁶⁹ Contextualists apply the idea of wide reflective equilibrium to the practical ethical sphere. They use the term "wide reflective equilibrium" to describe the process of achieving an equilibrium between possible judgements, theoretical considerations and background theories that bear on the situation. In the process of seeking equilibrium, none of the parameters bearing on the problem is foundational, all are open to revision. Solutions that best balance the multiple constraints are candidates for the final outcome.

In order to balance the various constraints acting on ethical problems it is useful to classify them. They fall into five basic categories: 1, possible judgements and outcomes; 2, values; 3, moral principles; 4, background theories; and 5, situational factors. Now, consider each in turn:

1. Possible judgements and outcomes: Possible judgements and outcomes of the problem are considered. For example, in a method called comparative case analysis in medical ethics, analogous situations and their resolutions are considered. The following situation will serve as an example:

A man with multiple sclerosis is admitted to the hospital for treatment of spinal meningitis with bacterial origin. His past history indicates a very satisfactory adjustment to MS. He has taken an active part in family life, he has had various interests, hobbies and so forth. Despite all of this, and despite the fact that his MS has not worsened, the man refuses antibiotic therapy to treat his meningitis, saying only that he wants to be left alone and allowed to die with dignity. Suppose consultation with the family reveals that the patient has been very withdrawn and depressed lately. It is also learned that the patient has been deprived for some time of the usual attentions and support of other family members because of a prolonged crisis elsewhere in the family. More evidence of the same kind makes it

¹⁶⁹ John Rawls [1971], 20f. See §2.6 for a more detailed discussion of Rawls.

fairly probable that the patient's decision is a product of a sense of self-pity and worthlessness accompanying feelings of isolation and depression.¹⁷⁰

Using comparative case analysis, this case would be considered in relation to cases relevantly similar to it. From this process of considering possible judgements, some of which may come from actual cases, one judgement may be chosen. At the same time, a new outcome or judgement may arise from the reflective equilibrium process. The various judgements and outcomes considered do not give the only possible judgements and outcomes for the situation, but rather provide a starting point for comparison and reflection.

Following the "paradigm theory"¹⁷¹ of bioethics, the case above would require that the physician respect the patient's autonomy and comply with his wishes. However, it seems clear that, in this particular case, alternative courses of action might be justifiable. An actual outcome is described in the following passage:

Physicians explain to the patient what they think is happening with him, they inform him decisively that they intend to give him antibiotics to save his life and that family counseling will be provided in due course. The patient is silent. Antibiotics are administered, the man recovers completely, family counseling reveals to the family the importance of this patient's being informed and involved in family affairs, and everything turns out well.¹⁷²

In a future comparative case analysis, the outcome suggested by the paradigm theory and the outcome given here can be examined in relation to the new situation. A more finely tuned response can be obtained when possible outcomes and judgements are compared and assessed in the wide reflective equilibrium process, than when a particular principle or theory is automatically applied to all cases.

¹⁷⁰ Winkler [1996], 65-66.

¹⁷¹ The "paradigm theory" is one that relies on three general principles: autonomy, beneficence (including non-maleficence), and justice. See Winkler [1996], 51.

¹⁷² Winkler [1996], 65-66.

2. Values: Beliefs arising from our framework of values are also balanced in the wide reflective equilibrium process. Each framework of values consists of values and norms which have accrued over the history of human social experience, which are often considered to be values and norms of common sense, and values related to one's understanding of oneself, as situated in the natural and the social world. One's self-concept can be revised and is often revised in light of considering various moral theories. In fact, Moody-Adams argues, one function of ethical theory and moral reflection is to help us reconsider our selves and our relation to the world in a new way.¹⁷³

Consider the examples Moody-Adams gives of how differences in self-conceptions can affect moral judgements. She argues that a person who has a self-concept involving possession of an immortal soul and a person who has a materialistic self-conception, may have a difficult time agreeing on the moral status of abortion.¹⁷⁴ Such disparate self-conceptions can become marked through moral interaction with others and then attention can be turned to whether or not to modify one's self-conception. This is an example of a situation in which the process of wide reflective equilibrium can be used to understand the disagreement on a particular judgement in terms of the underlying theoretical conflicts upon which they may be based. Daniels writes that the process of wide reflective equilibrium

may lead us to understand better the sources of moral agreement and disagreement and the constraints on what we count as relevant and important to the revision of moral judgments. It may allow us to reduce moral disagreements (about principles or judgments) to more resolvable disagreements in the relevant background theories.¹⁷⁵

Understanding and clarifying one's framework of values can be the result of thought experimentation. Our framework of values is both a ground for our intuitions, or

¹⁷³ Moody-Adams [1990], 230.

¹⁷⁴ Moody-Adams [1990], 231.

¹⁷⁵ Daniels [1979], 274.

moral perceptions and that out of which we further forge and develop our sense of values.

In this context, Nussbaum writes that moral perception is not a perceiving of the

world apart from human beings and human conceptual schemes, but a world already interpreted and humanized by our faculties and concepts.¹⁷⁶

That is, we interpret, measure and develop values through our existing framework of values.

There are two steps in the process of considering our values in relation to the thought experiment. First, we look to our own responses to a particular thought experiment. Second, we check our responses against those of others in order to assess the level of intersubjective agreement. The intersubjective process is important, because through it values can be clarified and developed. We see, then, two important features in the relationship between the thought experiment and the framework of values:

A. Responses to thought experiments arise from each person's framework of values.

B. A framework of values is itself developed in response to moral reflection, alone and with others, such as that afforded by the thought experiment.

Frameworks of values are revisable.

Thomson writes of the role of responses arising from a framework of values in the following way:

But which acts *are* the right one? Which acts *are* the wrong ones? Well, among the right ones are the ones which strike us as clearly, plainly, on any plausible moral view, right; and so also for the wrong ones.¹⁷⁷

¹⁷⁶ Nussbaum [1989], 131.

¹⁷⁷ Thomson [1986], 257.

I agree that we need to use what we have, our beliefs about what is right and wrong, as a starting point. From this point, we use wide reflective equilibrium to forge our positions. The following passage presents a clear description of this process:

We all begin, rightly or wrongly, with a haphazard collection of normative beliefs: lying is immoral, kindness is a virtue, it is wrong to torture innocent creatures simply for fun. Coming upon hard cases, we look to our previous storehouse of beliefs in the hope that we will find guidance. In doing so, we begin looking for some unifying principles, noticing, perhaps, that most of our previously held beliefs somehow serve to maximize human well-being, or happiness, or justice. After discovering such principles, we not only apply them to the troublesome hard cases under consideration. More than likely, we will also modify some members of our initial belief set in order to bring our previous beliefs more fully into line with these newly discovered principles. For example, in order to help maximize the coherence of our overall belief set, we may modify our original belief that lying is immoral, accepting in its place the belief that lying is immoral unless it aids in the promotion of some greater good. Spreading misinformation to the enemy in wartime, for example, may now be viewed as a moral, rather than as an immoral act. Such developments, in turn, may cause us once again to modify our newly discovered principles. In this way we will continue to try to maximize coherence, stopping only when we reach some satisfactory degree of reflective equilibrium.¹⁷⁸

In the following passage, Thomson suggests the use to which beliefs about values elicited through thought experiments can be put in the process of reaching a satisfactory solution about a moral question or issue:

[T]he moral theorist must attend to his or her own moral beliefs about examples, stories, and cases, actual or invented, looking to see how those beliefs do (or do not) change as the details of the stories are altered. For it is precisely those beliefs which supply the data for moral theorizing, and which go a long way — if not all the way — to setting the constraints on what constitutes an acceptable moral principle, and thus on what constitutes an acceptable way of understanding what we ourselves take morality to require of us.¹⁷⁹

A criticism of thought experiments concerning ethics pursues the incoherence of ordinary moral concepts. An account that appeals to frameworks of values, like mine, must answer this charge. The idea is that "[i]f common sense only aims at the prediction

¹⁷⁸ Irvine [1998], 562-563.

¹⁷⁹ Thomson [1986], 260.

and control of daily affairs rather than at describing reality, then folk concepts will behave erratically when pulled out of their workaday roles."¹⁸⁰ However, the erratic behaviour of concepts when pulled out of their workaday roles in thought experiments serves as a way to help clarify and develop more consistent moral concepts. Our thought experiments are purposefully designed to help bring out the very erratic qualities that cause some to want to avoid the enterprise entirely. Moreover, if thought experiments are going to have any relation to practice and any effect on practice they have to explore these terms as they are used and understood. This further benefit of using beliefs about values in an approach to ethical theorizing is noted by Moody-Adams. She writes,

[E]thical theory is pointless if it is not addressed to persons whose behaviour it aims to influence. It *must* start from the "inside": from the pre-theoretical deliverances of the moral consciousness of those to whom the theory is addressed.¹⁸¹

Moody-Adams argues against the objections which claim that an account of moral theorizing based on beliefs arising from a framework of values is nothing more than either an account of moral psychology or an account of naive acceptance of pre-theoretical convictions. Rightly, she emphasizes that it is not beliefs alone that are considered in moral reflection. In my account of wide reflective equilibrium we see this. Personal beliefs about values are balanced against a variety of constraints acting on the problem. In this process, the beliefs are subject to revision. Frameworks of value are brought under scrutiny when something challenges them and when they are seen to be inconsistent or inadequate.¹⁸²

Thought experiments are conducted against a framework of values that is internal to each experimenter. We share these internal experiments about ethics with one another and check on the varying reactions to them in order to calibrate our framework of values

¹⁸⁰ Sorensen [1992a], 18.

¹⁸¹ Moody-Adams [1990], 232.

¹⁸² Moody-Adams [1990], 232.

against those of others. These experiments in ethics are experiments that work on our beliefs about values directly.

Although the thought experiment works directly on our beliefs about our values through introspection about imaginary events, the most direct experiments about values are conducted in the fields of experimental psychology and economics. These experiments do not test what we think about our values, but rather what our values are as indicated by how we act and choose in actual situations.¹⁸³ These experiments and their results raise an important problem that is leveled against thought experiments in ethics. It goes like this: thought experiments may appeal to frameworks of values but we cannot trust the resulting responses. Values are not accessible on introspection. People deceive themselves. They embellish their responses and in general report what they think their values are rather than what in fact their values are.¹⁸⁴ Results of experiments conducted in the fields of psychology and economics show how it is people actually reason and respond to moral situations, and the results are not the same as those at which one might arrive by introspection. However, consider the following passage:

[T]wo questions about experiments with human subjects always need to be asked. Does the behaviour survive when the incentives are increased? Does it survive after the subjects have had a long time to familiarize themselves with all the wrinkles of the unusual situation in which the experimenter has placed them? If not, then the experimenter has probably done no more than trigger a response in the subjects that is adapted to some real-life situation, but which bears only a superficial resemblance to the problem the subjects are really facing in the laboratory.¹⁸⁵

What we see here is that there is a distance between these actual experiments and "real life", just as there is a difference between thought experiments and "real life". To examine beliefs about values and to examine human behaviour are two different things. Thought experiments explore the beliefs about our values. However, our cognitive

¹⁸³ See Skyrms [1996] for a surveys of these experiments.

¹⁸⁴ Baier [1987], 49.

¹⁸⁵ Binmore [1994], 185.

processes can affect our frameworks of values which later may affect behaviour. Moreover, in considering our values as indicated by our responses, we can weigh them against evidence of our values which we have gained in examining our responses to actually experienced situations, and experimental results from psychology and economics. In this way we can check whether our beliefs about our values match our values as they are indicated by our behaviour. In Chapter 7, I examine ways in which we can elicit the responses to thought experiments which are more likely to reflect the values that inform our decisions and actions in everyday life.

3. Moral Principles: In describing the process of wide reflective equilibrium, Winkler explains how we develop a set of principles which brings considered judgements into coherent order. He writes:

The process of theory building can be said to begin with our most secure considered judgements, which may be either general or particular. Next we develop a set of principles which rationally orders and explains these judgements. This is the first criterion for judging the acceptability of a set of principles, their ability to bring the whole array of our considered judgements into coherent order.¹⁸⁶

Principles are a way of generalizing particular moral judgements. We justify principles by considering their explanatory abilities. Explanatory ability is judged by the manner in which the principles order our judgements arising from within our framework of values. For example, in the MS case given above, we may feel strongly that the MS patient ought to be treated. If this is the case, we will need to see what moral principle can be evoked to explain this reaction. The principle we might consider is that intervention against the patient's wishes may be attempted only at times when we have evidence that psychological factors, such as temporary, or treatable depression are affecting the patient's reasoning about his or her future.

¹⁸⁶ Winkler [1996], 64.

4. Background Theories: In the wide reflective equilibrium process, background theories that are pertinent to the problem will be considered. For example, in considering a problem in medical ethics, such as that of the MS patient, we need to consider psychology, medicine and the operation of healthcare facilities. At a deeper level, epistemological and logical theories that deal with assumptions regarding the sources of knowledge claims and their relative weighting form part of the background to the moral problem.

5. Situational Factors: Situational factors that make the ethical problem unique are also considered. In the MS case the situational factors regarding the family, the reasons for the onset of depression and the prior high-functioning of the individual all bear on the situation and should be considered in order to obtain the most appropriate solution.

Each contextualist process is done in relation to actual practices, and so the equilibrium struck between possible judgements, beliefs about values, principles, background theories, and situational factors applies in a particular context. What is correct in one context is a function of a complex set of conditions that affect and characterize that particular situation.¹⁸⁷

Rawls' idea of reflective equilibrium is used to suggest a way to capture a particular community's idea of justice by bringing their ideas about specific cases into equilibrium with their principles. In ethics we can see that this idea has appeal. We have specific ideas that are embedded in our frameworks of values. Then we have broader principles also embedded in our frameworks of values. Thought experiments can aid in bringing our considered judgements about particular cases into equilibrium with our ethical principles and other relevant theories.

¹⁸⁷ Winkler [1996], 52-53.

The difference between Rawls' reflective equilibrium and the wide reflective equilibrium discussed here is the absence of a veil of ignorance. The negotiators for the principles of justice which are to govern the society of the people whom the negotiators represent, do not know anything about the people whom they represent. The idea of the veil of ignorance is that it excludes that knowledge about the individual participants in the future society which could bias the negotiators. The only knowledge the negotiators have is knowledge that is relevant to considerations of justice. The knowledge they do not have is that which is generally agreed to be irrelevant to considerations of justice. By contrast, in the wide reflective equilibrium process in ethics, the intent is to include as many of the contextual features of the ethical situations as possible in order to achieve a result that is relevant to the context to which it will apply. There is no veil of ignorance, since, on the contextualist account, all details of the situation are considered as possibly relevant to the ethical process.

We can contrast this contextualist method of problem solving with that arising from a more traditional, foundationalist conception of ethics. On the foundationalist model, a problem is solved by reference to a set of normative principles. These normative principles arise from a theory of human nature or metaphysics that is used to justify judgements in all ethical situations. In order to solve a particular problem, there is an appeal to a rule which comes from an underlying moral principle set within a foundational normative theory.¹⁸⁸

A useful analogy can be drawn between foundationalist and coherentist theories in epistemology and foundationalist and contextualist views in ethics. In foundationalist epistemology, knowledge is justified by its relation to a foundation of certain truths. In traditional foundationalist ethics, a judgement is justified by its relation to the moral principles derived from a basic normative theory. In the coherentist model, knowledge is

¹⁸⁸ Winkler [1996], 50.

justified in relation to its coherence within a web of beliefs. In contextualist ethics, a judgement is justified if it best balances in a coherent manner the various constraints or parameters bearing on the situation.

4.2 The Design Problem Approach

Coming to an adequate solution for a moral problem can be compared to finding a solution to a problem in applied science. Caroline Whitbeck writes about the similarities between ethical problems and engineering design problems. Rather than thinking of ethical problems as dilemmas with a number of pre-prescribed conflicting solutions, she suggests that we consider them as design problems with a number of solutions, each of which will balance the various requirements for resolution. Of these various solutions she holds that there is rarely only one uniquely correct solution, even though some solutions may be better than others. Moreover, in comparing any two solutions, each one may have a different set of advantages over the other.¹⁸⁹ The possible solutions must conform to criteria and specifications for the solution to the problem, and must not be in conflict with various background considerations.¹⁹⁰

Design problems are not dilemmas but, rather, problems of "making (or repairing) things and processes to satisfy wants and needs".¹⁹¹ Like ethical problems, design problems are often highly constrained. When ethical problems are not seen as dilemmas, but rather as design problems, the focus moves from trying to choose from two mutually exclusive paths of action to trying to understand the problem clearly and devise a number of solutions that will satisfy it.¹⁹² Whitbeck writes,

¹⁸⁹ Whitbeck [1996], 11.

¹⁹⁰ Whitbeck [1996], 12.

¹⁹¹ Whitbeck [1996], 10.

¹⁹² Whitbeck [1996], 9.

Many moral problems that are represented as conflicts are better understood as problems in which there are multiple constraints that may or may not turn out to be simultaneously satisfiable.¹⁹³

The ideas given by Whitbeck have their parallel in the wide reflective equilibrium process for solving ethical problems. Just as in the wide reflective equilibrium process one considers possible judgements, outcomes, values, moral principles, background theories and situational factors, so in designing an artifact, or a process, one must consider possible outcomes, values, physical principles, background theories and situational factors. For example, although there is no uniquely correct solution to how best to build a bridge, in designing a bridge an architect will consider various possible bridge designs; values, such as the prevailing styles and attitudes toward architecture, and those arising from the history and culture in the area; the principles of physics; background theories such as climatic effects on bridge structure, slope stability theory, tides and current predictions; as well as situational factors such as traffic flow, and financial restraints.

The design approach can be characterized as one employing the process of wide reflective equilibrium in order to satisfy the requirements and constraints of a problem based on the needs and desires of those involved. Winkler's view of contextualism is consistent with the design approach. Winkler writes,

We can think of social moralities as social instruments for the promotion and maintenance of valuable forms of life. The moral point of view is now essentially focussed on how we ought to relate to one another in the service of this general end, as it is embodied in the various primary domains of social life.¹⁹⁴

In the process of applying wide reflective equilibrium to ethical problems, frameworks of values develop. Each new solution to an ethical problem has an effect on this development. By analogy, consider the experiments that led to the design of the

¹⁹³ Whitbeck [1996], 15.

¹⁹⁴ Winkler [1996], 58.

steam engine. Here, we see the interaction between the design process and theory development. Hacking presents the view that these experiments began as practical inventions that led to the development of theory. He writes,

One road to new technology is the elaboration of theory and experiment which is then applied to practical problems. But there is another road, in which the inventions proceed at their own pace and theory spins off on the side. The most obvious example is the best one: the steam engine.¹⁹⁵

Thought experiments in ethics can be viewed as inventions in this way. We construct a mental model of a situation and a response. The comparison of cases is conducted in relation to the practical context of particular responses, which later get codified into values and moral principles which form part of a framework of values.

Hacking writes about the progress of the steam engine technology viewed as a process of experiment. He writes:

The experiments were the imaginative trials required for the perfection of the technology that lies at the centre of the industrial revolution.¹⁹⁶

In the same way, thought experiments in ethics can be seen as imaginative trials leading to more developed responses to ethical problems, more consistent values and refined moral principles.

4.3 A Contextualist Approach to Thought Experiments in Ethics

Judith Jarvis Thomson presents the following thought experiment about "people-seeds" in order to argue that abortion is morally permissible, particularly in a situation where precautions against pregnancy have been taken.

If the room is stuffy, and I therefore open a window to air it, and a burglar climbs in, it would be absurd to say, 'Ah, now he can stay, she's given him a right to the use of her house — for she is partially responsible for his presence there, having voluntarily done what enabled him to get in, in full knowledge that there are such things as burglars, and that burglars burgle.'

¹⁹⁵ Hacking [1983], 163.

¹⁹⁶ Hacking [1983], 164.

It would still be more absurd to say this if I had had bars installed outside my windows, precisely to prevent burglars from getting in, and a burglar got in only because of a defect in the bars. It remains equally absurd if we imagine it is not a burglar who climbs in, but an innocent person who blunders or falls in. Again, suppose it were like this: people-seeds drift about in the air like pollen, and if you open your windows, one may drift in and take root in your carpets or upholstery. You don't want children, so you fix up your windows with fine mesh screens, the very best you can buy. As can happen, one of the screens is defective; and a seed drifts in and takes root. Does the person-plant who now develops have a right to the use of your house? Surely not — despite the fact that you voluntarily opened your windows, you knowingly kept carpets and upholstered furniture, and you knew that screens were sometimes defective. Someone may argue that you are responsible for its rooting, that it does have a right to your house, because after all you *could* have lived out your life with bare floors and furniture, or with sealed windows and doors. But this won't do — for by the same token anyone can avoid a pregnancy due to rape by having a hysterectomy, or anyway by never leaving home without a (reliable!) army.¹⁹⁷

Thomson compares our excusing of responsibility toward a person whose house has been robbed, even though she, or he, has taken all appropriate precautions, with our lack of excusing responsibility toward a woman who has become pregnant, even though she too has taken all appropriate precautions. She shows how, if we are to be consistent, we should extend our responsibility and blame ascriptions in the same manner to pregnant women as we do to robbed women.

The contextualist process of wide reflective equilibrium is useful in considering thought experiments in ethics, in that it allows us to consider the relationship of the thought experiment to all the factors which make up its theoretical context. Daniels notes that it is a virtue of wide reflective equilibrium that it incorporates a way to access the various considerations people give as reasons or evidence in moral arguments and weigh them in a way that is relevant to underlying theories.¹⁹⁸ In creating a thought experiment, grounding will be facilitated through employing the wide reflective equilibrium process which weighs possible judgements and outcomes, beliefs arising from the framework of

¹⁹⁷ Thomson [1971], 48-49.

¹⁹⁸ Daniels [1979], 262.

values, moral principles, background theories and situational factors. Wide reflective equilibrium can also help us assess the extent to which a thought experiment is grounded. I use this latter function of wide reflective equilibrium in the following examination of Thomson's people-seed thought experiment.

In Thomson's thought experiment, the possible outcomes and judgements are presented for us. We are asked to decide if people-seeds have an absolute right to live in the house and by analogy whether a foetus has an absolute right to life. Thomson argues that just as people-seeds do not have a right to live uninvited in the house, the foetus does not have an absolute right to life, unless we are willing to say that bare floors and hysterectomies are mandatory.

The primary theory in this thought experiment is our framework of values. Notice that I use the term "framework" rather than "theory" in referring to the set of beliefs that express our values. This set of beliefs is largely unarticulated, often inconsistent and has comparatively few of the attributes of a carefully thought out and developed theory. It is partly due to this feature of the primary background "theory" of thought experiments concerning ethics that they are so difficult to ground. To the extent that the Thomson thought experiment is grounded, it is because the propositions it contains conform to our framework of values. Thomson compares her new case about people-seeds to the robbery case. We have an accepted solution to the robbery case, which we feel applies to the people-seeds case also. Thomson then compares the people-seeds case to that of unwanted pregnancy, in spite of precautions taken against it. We may not have an accepted solution to the pregnancy case in our particular framework of values, but we are led to consider applying our judgement in the case of people-seeds and robbery to the unwanted pregnancy case.

We then consider how our beliefs from the framework of values can be ordered into moral principles. We see that we do not generally ascribe responsibility for an event to someone who has tried, as far as is reasonably possible, not to cause that event.

In order to decide which background theories affect the outcome of a thought experiment, we need to examine those propositions which describe the concepts used in the experiment and also those propositions which describe the relationships between concepts in the thought experiment. These propositions will indicate a certain set of theories that make up the theoretical context of the experiment.

Consider the concept of responsibility as used in Thomson's thought experiment. When we look to sentences describing or using this concept, we get sentences such as:

1. Being responsible for something in some way is having to answer for that thing in that way.
2. Responsibility can be self-imposed or other-imposed.
3. Responsible people usually have some measure of control over the things and circumstances for which they are responsible.

None of these sentences are necessarily about the physical world and so even with this cursory examination we might expect that an experiment focusing on responsibility need not be grounded in physical laws. The thought experiment will not have to conform to physical theory since here the goal is to expose inconsistencies in our framework of values. The counterfactual situation involving people-seeds is analogous to the unwanted pregnancy situation with respect to this framework. Physical theory, in this case, is independent of our framework of values. This may not be the case in other situations. For example, certain parts of physical theory and a particular framework of values may not be independent in questions regarding problems in biomedical ethics such as euthanasia. In this case, however, physical theory and the framework of values are independent.

By contrast, consider the background theory of a thought experiment about personal identity. In order to do this, consider some of the propositions describing the nature of personal identity:

1. Personal identity is a feature of individual human beings.
2. Personal identity is both self-attributed and other-attributed.
3. Personal identity is connected with the psychological or physical continuation of a person.

The second sentence invokes social theory. The third sentence invokes both psychological and physical theories. It thus seems reasonable to conclude that personal identity theory, social theory, psychological theory and physical theory are interdependent. When conducting thought experiments about personal identity, we must therefore use propositions that are relevantly related to these background theories (except for those propositions that describe extraneous particulars).

We have seen that Thomson's thought experiment about responsibility does not have to be set in physical theory. However, the ethical terms will need to conform to our beliefs arising from our framework of values. In Thomson's thought experiment, there are no values presupposed that do not accord with standards that we understand and accept from our common, historically embedded moral values. Thomson is exploring well understood moral precepts in order to clarify their scope of application.

Finally, we consider any unique situational aspects that might affect the outcome of the experiment. We see that in this case there are not any. The experiment may seem to be situationally unique but it is an analogy for a fairly ordinary event, a case of unwanted pregnancy in spite of precautions taken against it. We see that there are no situational factors that cause us to override the general moral principles and beliefs about responsibility arising from our framework of values.

In the last chapter, we saw that grounding concerns the relationship of the thought experiment to its theoretical context. The concepts of which our theories are composed arise out of the data of empirical observation and beliefs arising from within our framework of values. Concept formation is mediated by philosophical, social, and scientific conventions and background theories. In order to have thought experiments that are grounded, the concepts must be defined normally and the propositions of the thought experiment (except those relating to extraneous particulars) must be relevantly related to the theoretical context. In Thomson's people-seed thought experiment, concepts are used in their ordinary way and the propositions are relevantly related to the theoretical context.

In ethics, the thought experiment is formed and responded to in relation to a framework of values. In order to see if any other factors from the theoretical context affect the grounding it is useful to apply a consideration of the various factors that influence the thought experiment, just as we do in the wide reflective equilibrium process. Possible judgements and outcomes, beliefs about values, moral principles, background theories and situational factors are considered. In the process of seeing how the thought experiment relates to these factors we can see if the thought experiment is relevantly related to the theoretical context in the manner outlined in Chapter 3 and so determine whether it is grounded or not.

For the purposes here, the process of wide reflective equilibrium is not used in support of a particular moral or ethical position. Rather, it is used as a tool with which we can ascertain whether the thought experiment is grounded or not. Since thought experiments in ethics are conducted against a variety of primary and background theories, it is appropriate to have a way to check for the grounding of these thought experiments that does not invoke a particular ethical theory. Wide reflective equilibrium can be used in this function, as a means to check the relationship of the thought experiment to the particular theory and the particular frameworks of values that are involved.

We will come back to thought experiments concerning ethics in Chapters 6 and 7, but in the next chapter it is instructive to turn to thought experiments in other areas and see how an analysis of them relates to the account given so far.

5. Grounding Thought Experiments in Various Contexts

In the last chapter, I examined the grounding relationship for thought experiments in ethics. Here, I consider the grounding relationship with respect to a number of other theoretical contexts.

Irvine writes of scientific thought experiments that

[m]any, although not all, of the assumptions within the thought experiment must be supported by independently confirmed empirical observation. In short, at least some of the features of the thought experiment must be grounded in the observable world if it is to have any relevancy to general scientific inquiry.¹⁹⁹

The data for ethical theory come mainly from our framework of values and are mediated mainly by social and philosophical conventions. In contrast, the data for physical theory come mainly from empirical observation and are mediated mainly by scientific and theoretical conventions. Since the domain of scientific thought experiments is the physical world, these thought experiments must be grounded in physical theory which is then grounded in independently confirmed empirical observation.

Here I consider how my account of grounding applies to thought experiments regarding not only science, but also philosophy of language and personal identity theory. This examination serves to highlight some of the points I made about grounding in Chapter 3. In particular, we see how features of a specific field affect the grounding relationship. For example, some thought experiments, such as those in the philosophy of personal identity, rely on a number of interconnecting theories; while some thought experiments, such as those in physics, rely on a relatively well-defined and circumscribed theory. Consideration of these cases helps us consider how features of the theoretical context of thought experiments in ethics affect the grounding relationship.

¹⁹⁹ Irvine [1991], 159.

5.1 Thought Experiments in Philosophy of Language

I include a thought experiment about philosophy of language to illustrate the only type of situation in which my grounding condition regarding concepts that are defined normally does not apply. Recall that concepts must be defined normally, except when it is the concept itself that is under scrutiny. In the following thought experiment, the dependent variable concerns the truth of the conjunction of two assumptions derived from the Fregean theory of meaning. The independent variable is the way that "water" relates to narrow psychological states and the world. I stated in Chapter 3 that the independent variable must conform to the propositions which express the concepts it employs, *unless* the concept itself is under scrutiny. This type of thought experiment can occur with respect to any theoretical context. In the case given here, the independent variable does not conform to the propositions which express the concept of water as it is defined normally.

Thought experiments in philosophy of language often explore the use and meaning of concepts. In *Meaning and Reference*,²⁰⁰ Putnam introduces a thought experiment about the use of the word "water" to show that meanings are not in the head. That is, Putnam's thought experiment is designed to clarify the relationships among meaning, psychological states and extension.

Suppose that there is a planet called Twin Earth that is exactly like Earth except for a few differences. One difference is that what the people on Twin Earth call "water" is not H₂O, but rather a liquid whose chemical formula we will call XYZ. XYZ is indistinguishable from water, in most ways. On Twin Earth, the lakes and oceans contain XYZ, it rains XYZ, XYZ is good to drink, XYZ freezes and boils as water does, and so on.

²⁰⁰ Putnam [1973].

An Earthian space ship visiting Twin Earth and discovering that "water" on Twin Earth is XYZ would report:

On Twin Earth "water" means XYZ.

Similarly, a Twin Earthian spaceship, upon discovering that "water" on Earth is H₂O would report:

On Earth "water" means H₂O.

Now consider Twin Earth and Earth circa 1750. At this time, the Twin Earthian did not know that "water" consists of XYZ and the Earthian did not know that water consists of H₂O. Oscar₁ lives on Earth and Oscar₂ on Twin Earth. Oscar₁ and Oscar₂ are exact duplicates in feelings, thoughts, appearances, etc. At the same times and in the same surroundings they do and say the same things. There is no belief that Oscar₁ has about water, that Oscar₂ does not have about "water". That is, Oscar₁ and Oscar₂ are in the same narrow psychological state. Yet, the extension of "water" on Earth is H₂O and the extension of "water" on Twin Earth is XYZ. We can conclude that Oscar₁ and Oscar₂ understand the word "water" differently even though they are in the same narrow psychological state. This would be true even though their respective scientific communities had not yet discovered the chemical composition of what the term "water" referred to on each planet.

Putnam's goal is to offer a reverse example to the one in which the two terms "creature with a heart" and "creature with a kidney" are given as an example of two terms that have the same extension but differ in meaning. His Twin Earth thought experiment is designed to show that two people can be in the same narrow psychological state with respect to a term while the term differs in extension for each person. Putnam wishes to challenge the theory of meaning coming from a Fregean view of what it means to

understand an expression. This theory of meaning rests on the following two assumptions, which, Putnam claims, are not jointly satisfied by any notion of meaning:

Assumption I: That knowing the meaning of a term is just a matter of being in a certain psychological state.

Assumption II: That the meaning of a term determines its extension (in the sense that sameness of intension entails sameness of extension).²⁰¹

The first part of the Twin Earth argument is designed to show that the two assumptions cannot be true together. Oscar₁ and Oscar₂ have the same narrow psychological states, however for Oscar₁ "water" has as its referent H₂O, while for Oscar₂ "water" has as its referent XYZ. With respect to assumption I, both Oscar₁ and Oscar₂ know the meaning of "water" as it is employed in their respective linguistic communities, and are in the same narrow psychological states. They have the same narrow beliefs about water. With respect to assumption II, there is no difference in Oscar₁'s and Oscar₂'s narrow psychological states; they have the same beliefs about the sense of the word "water", but there is a difference in the extensions of the term "water". Either the meaning of a term does not entail being in a certain psychological state (Assumption I is false), or the meaning of a term, the knowing of which is a matter of being in a certain psychological state, does not determine its extension (Assumption II is false). Putnam concludes that "Meanings ain't in the head."²⁰²

Putnam concludes that meaning is connected with extension. He supports this conclusion with the second part of the thought experiment, about Earth and Twin Earth circa 1750. This part of the thought experiment is designed to show that meanings are not in the head in the following way. When Oscar₁ points to water and says, "This is

²⁰¹ Putnam [1973], 308-309.

²⁰² Putnam [1973], 311. Later in the paper Putnam considers what does determine the extension of a term.

water", he is picking out what he is pointing to, and all other things like it, and labeling them as water. When Oscar₂ points to water and says, "This is water", he is picking out what he is pointing to, and all other things like it, and labeling them as water. With this ostensive definition he is saying that the stuff he points to is called "water" and is the same as the stuff that others in his linguistic community call "water". "Water" for Oscar₁ picks out H₂O in every possible world, while "water" for Oscar₂ picks out XYZ in every possible world. Here again meanings are not in the head but rather have something to do with what is picked out by ostension in the external world. It is the difference in the worlds that makes the difference in extensions. Water and other natural kinds terms, Putnam holds, are rigid designators. A designator is rigid if it designates the same thing in every possible world in which that thing exists. It follows that "Water is H₂O" is a necessary truth since these terms pick out the same thing in every possible world in which there is water. However, this sentence is not epistemically necessary since it was rational for people to believe that water is not H₂O before 1720, and it may later be rational for us to believe it is not H₂O but rather H₂O plus a small amount of XYZ.

The conditional that shows how the variables relate for this thought experiment is as follows:

If "water" has the same narrow psychological state for two people, even though it refers to different substances, then the theory of meaning that claims that Assumption I and II are jointly true is false.

Here, the primary theory concerns the conceptual realm rather than the empirical world. What is in question in the experiment is how our concept of water relates to its extension and to psychological states. The independent variable varies the extension for the term "water" while holding the narrow psychological state steady. When we consider that "water" has the same narrow psychological state for both Oscars regardless of its extension we see that meanings are not in the head. The use of counter-factuals regarding

Twin Earth and Twin Oscar does not affect the thought experiment because the independent variable is varied in a way that is relevantly related to the primary theory (the theory of language) and the question under examination (how our concept of water relates to its extension and to psychological states).

In Chapter 3, I claimed that the concepts used in thought experiments must be defined normally, unless it is the concept itself which is under scrutiny. The twin earth thought experiment is a case where the concepts themselves are under scrutiny. In most thought experiments we want the propositions to use concepts that are defined normally. That is, they must be employed in a way that reflects everyday usage and understanding based on the theoretical context. When a concept is defined normally, it is characterized by propositions which are consistent, plausible and testable with respect to the theoretical context. Here, however, we do vary the propositions which comprise the concepts in terms of the extension and psychological states, since these propositions are related to the relationship of the dependent and independent variables.

In this thought experiment, the purpose is to test how the meanings of concepts and their referents relate. We take a concept, "water", and test it in a situation in which the world is different in order to understand how it is that our understanding, concepts and the world link up. This thought experiment is grounded because it compares "water" as it is used and understood, in our world, against an analogous situation in which only things affecting the variables in question are altered in order to test the original concept. The statements conform to the theoretical context, except those that relate to the independent variable, the concept of water. In thought experiments where the independent and dependent variable relate to a question about the meaning, applicability or boundaries of a concept, the concept does not need to be defined normally. However, the concept must be varied in a way that is relevantly related to the question under examination.

5.2 Thought Experiments Regarding Personal Identity

The concept of personal identity has physical, biological, psychological, logical and social components. How these are to be balanced is a question of debate. Many personal identity thought experiments are designed to clarify these very problems. I consider the grounding of thought experiments concerning personal identity in order to show the difficulty of ascertaining the grounding of thought experiments in which the theoretical context is unclear.

Before considering a personal identity thought experiment, consider a well-known thought experiment about identity itself — the ship of Theseus thought experiment:

The vessel in which Theseus sailed and returned safe, with those young men, went with thirty oars. It was preserved by the Athenians to the times of Demetrius Phalerus; being so pieced and new framed with strong plank, that it afforded an example to philosophers, in their disputations concerning the identity of things that are changed by growth; some contending that it was the same and others that it was not.²⁰³

Hobbes' addition to the ship of Theseus develops and sharpens the question of identity:

For if, for example, that ship of Theseus, concerning the difference whereof made by continual reparation in taking out the old planks and putting in new, the sophisters of Athens were wont to dispute, were, after all the planks were changed, the same numerical ship it was at the beginning; and if some man had kept the old planks as they were taken out, and by putting them afterwards together in the same order, had again made a ship of them, this, without doubt, had also been the same numerical ship with that which was at the beginning; and so there would have been two ships numerically the same, which is absurd.²⁰⁴

This thought experiment reveals that a concept of identity that is based on the idea of sameness over time is problematic. It presents a situation in which one ship has continuity of form with the original ship of Theseus, while the other has continuity of material. Since identity is a one-to-one relationship, both ships cannot be identical with the original.

²⁰³ Plutarch, 22-23 (*Life of Theseus*).

²⁰⁴ Hobbes [1655], 84-85 (*De Corpore*, 2, 11, 7).

With respect to grounding, the concepts used in the "ship of Theseus" are defined normally, and the propositions which describe the thought experiment are relevantly related to those of the theoretical context. There are no simplifying assumptions, or extraneous particulars. The independent variable (replacing and restoring the planks one by one), relates to both the question under exploration and the theoretical context.

The primary theory is that of identity. It is instructive to consider that there could be physical inconsistencies included in this experiment that are not relevant to the conclusion. The lumber could be "beamed" across the planet to a ship yard in another country, or the whole episode could have occurred on Alpha Centauri. In this experiment we are able to build and repair boats, but if we did not have that craft, we could still, without altering our ideas about boat identity or the identity of material objects in general, imagine that this ability would be theoretically possible. These propositions would not need to conform to the theoretical context because they relate to experimental conditions not affecting the variables.

Now consider a personal identity thought experiment.²⁰⁵ Imagine that a person, Smith, has the right hemisphere of his brain destroyed and the left hemisphere transplanted into a body with no brain. The new person, Lefty, is psychologically similar to Smith. (This supposition comes from studies of persons who have survived the loss of half a brain.) In this case, we are likely to conclude that Smith and Lefty are the same person. Now imagine that the right half of Smith's brain is not destroyed but rather transplanted into another body. The resulting person is Righty. Now we have the case where we must, if we are to be consistent, imagine that both Righty and Lefty are psychologically similar to Smith, and yet we cannot at the same time consistently say that

²⁰⁵ This version is given by Sorenson, [1992a], 11-12. For earlier versions of the "fission" thought experiment see Wiggins [1967], 52-53; and Shoemaker [1984], 84-5.

they are both Smith. This example raises problems with the idea of psychological continuity as a basis for establishing personal identity.

Now consider whether the Lefty/Righty experiment is grounded in its theoretical context. It seems to succeed in challenging our concept of psychological continuity. The general question Lefty/Righty explores is "What is personal identity?" The question is explored by comparing two situations:

1. If we transfer half of a brain to another body and destroy the other half of the brain, then the resulting person is the same as the original person, since that person is psychologically continuous with the original person.
2. If we transfer half of a brain to each of two bodies, then the resulting persons are the same as the original person, since these persons are psychologically continuous with the original person.

This thought experiment tests our criteria of psychological continuity for personal identity. The variables can be expressed in this conditional sentence:

If two people can be psychologically continuous with a previous person, then psychological continuity cannot be a criterion, or the sole criterion, for personal identity.

Here, the theoretical context is complex. The primary theory, personal identity theory, is related to several background theories, including:

1. Social theory: How do we attribute personal identity in social contexts?
2. Physical theory: How do we attribute personal identity in physical contexts?
3. Concept of identity (as opposed to our concept of personal identity which is what is under examination): What is our concept of identity more generally?

Lefty/Righty presents the following puzzle: we look to our custom of identifying a person who is psychologically continuous with another as constituting identity. However, we also hold that conceptually, two physically separate persons cannot be thought of as one. As in the "ship of Theseus", we see that the concept of identity based on the criterion of sameness of some particular feature over time is problematic.

Wilkes argues that the brain-splitting experiments are too far-fetched to be taken seriously.²⁰⁶ She claims that the experiment is not grounded at the physical (data) level because the brain-splitting scenarios ignore the fact that our concept of identity is supervenient on the physical reality of our present embodied state. For, although we have studied people who have lost half of their brain and have found that they are psychologically similar, it seems unlikely that the procedure described by Shoemaker is physically possible. We do have experience with hemispherectomies²⁰⁷ and commissurotomies,²⁰⁸ but neither of these procedures involve the removal of a entire half brain since both leave the subcortical regions of the brain untouched. These regions are crucial to all psychological functions.²⁰⁹ From this information, it seems unlikely that we can divide the brain in half, and transplant each half into a brainless body. It seems likely that this procedure would lead to one fewer, rather than one more, individual and that the psychological continuity of Smith would be terminated rather than split.

Yet, one might object to this claim that the lack of grounding in physical theory tells against the thought experiment. Why should one worry about the impossibility of this type of brain surgery when we do not worry about the impossibility of beaming lumber from one ship to another? Why are some factors relevant in considering

²⁰⁶ Wilkes [1988], 36.

²⁰⁷ Hemispherectomy is the surgical procedure that removes the cerebral cortex of one hemisphere of the brain.

²⁰⁸ Commissurotomy is the surgical bisection of parts of the corpus callosum, a band of fibres which connects the two halves of the cerebral cortex.

²⁰⁹ Wilkes [1998], 38.

grounding while others are not? The answer lies in considering the question the experiment is designed to elucidate, the dependent and independent variables that express the set-up of the experiment, and their relation to the theoretical context of the thought experiment itself. Then, we need to ask whether the independent variable is varied in a way that is relevantly related to the question and theoretical context.

I agree that our concept of personal identity is affected by the background theory of biology, but upon examination, we see that in the Lefty/Righty thought experiment the background theory is modified only in relation to the independent variable. Recall that the independent variable does not have to conform to the theoretical context of the thought experiment. Recall also that simplifying assumptions and facilitating extraneous particulars do not need to conform to the theoretical context.²¹⁰

Even though Wilkes may claim that the Lefty/Righty thought experiment fails, I argue that it does not. The Lefty/Righty thought experiment meets my criteria for being grounded and successfully challenges the criterion of psychological continuity for personal identity. The result is that we need to discard or modify our criterion of psychological continuity for personal identity. The experiment varies only the independent variable of the number of people who can be psychologically continuous with a previous person. In order to isolate this variable, Putnam employs improbable physical processes such as brain transplants, but these physical processes are extraneous particulars of the thought experiment. We are not worried about whether these processes are possible, we are worried about whether it can be the case that two people who are psychologically continuous with a previous person can be said to have the same identity. The experiment leads to the consideration that when we think about personal identity, it includes a whole cluster of concepts about what it means to be biologically determined, embodied, individual beings. In Kuhn's terminology, the experiment brings to light a

²¹⁰ See §3.3.2.

possible incompleteness in our concept of personal identity and so leads us to bring this concept under further scrutiny and to interpret it in a new way. Instead of holding that psychological continuity is a criterion for personal identity, one might now hold that the concept of human personal identity also needs, in its core, facts about what it is to be biologically human, and that these facts are governed by laws about what is and is not possible for human life. This position is in some respects similar to Wilkes' position. The difference is that Wilkes proclaims the position in spite of the thought experiment, while I see it as a result of the thought experiment's successful refutation of psychological continuity as a sole criterion for personal identity.

It is interesting to note that it seems difficult to decide whether Lefty/Righty is grounded because of inconsistencies in its theoretical context. Thought experiments such as this one expose inconsistencies in our developing theories. The thought experiment is a good example of the point that the clearer the theoretical context, the easier it is to tell whether an experiment is grounded. Despite this, there are good thought experiments in personal identity which clearly do manage to isolate their variables. These types of thought experiments help develop more well-defined theory. Consider Thomas Reid's thought experiment designed to refute Locke's claim that personal identity requires continuity of memory with one's past. He writes,

Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life; suppose also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.²¹¹

On Locke's account, continuity of memory is a condition for personal identity and so the old general and the young officer are the same person, and the young officer and the boy are the same person, but the boy and the general are not the same person. This thought

²¹¹ Reid [1785], 213-4 (Essay 3, Chapter 6).

experiment succeeds in calling for a rejection or modification of Locke's theory. Reid has given us a thought experiment which is grounded. The concepts are defined normally and its propositions are consistent with the theoretical context. The independent variable — continuity of memory — is varied in a way that is relevantly related to the question under consideration.

5.3 Thought Experiments in Science

In examining thought experiments in science we see that due to the relatively well-developed and well-defined theoretical context, the question of whether or not the thought experiment is grounded is usually easy to ascertain. To see this, consider two well-known thought experiments in science.

Einstein, at age sixteen, imagined moving as fast as light and so being able to travel on the front of a light beam. By imagining what would happen if one moved with the same speed as an object in our frame of reference, Einstein reasoned that he would see the light beam as a stationary oscillatory field. However, seeing a light beam as a stationary oscillatory field is inconsistent with Maxwell's theory of electrodynamics which states that an oscillating electromagnetic field is essential for a light wave. This thought experiment, and the inconsistency it exposed, played a role in the genesis of Einstein's theory of special relativity.²¹²

In this case, the light beam thought experiment is grounded in the primary theory which in turn is grounded in confirmed empirical observation. We are able to observe what things look like in our frame of reference when we are moving at the same speed — they look stationary — and we know that light is composed of an oscillating magnetic field. The light beam thought experiment involves impossibilities. However, the impossibility concerns one of the extraneous particulars of the experiments. The fact that

²¹² Given in Brown [1991a], 15-16.

one cannot run with a light beam is not relevant to the primary theory of the experiment since human running abilities are independent of Maxwell's theory of electrodynamics. Human capacities are not central to the questions of electrodynamics.²¹³

Second, consider the Stevin's chain thought experiment given in §2.4.2. Recall, Stevin concludes his experiment with the observation that

On inclined planes of equal heights weights act in inverse proportion of the lengths of the planes.²¹⁴

The concepts used in this thought experiment are defined normally and, except for the simplifying assumption, the propositions that describe the thought experiment conform to physical theory. The simplifying assumption allows the experimenter to consider the relationship between weight, height and velocity independent of the effect of friction.

In this chapter so far, we have seen how the account of grounding applies to thought experiments in philosophy of language, personal identity theory and science. The thought experiment in philosophy of language serves to illustrate when it is that a thought experiment does not have to employ concepts that are defined normally. The personal identity thought experiments show us how a developing and sometimes vague theoretical context makes grounding difficult. This is particularly true, when the theoretical context of the thought experiment is one composed mainly of common-sense beliefs.²¹⁵ By contrast, thought experiments concerning science are often more easily grounded, and it is easier to ascertain whether they are grounded, because their theoretical context is usually more well-defined.

²¹³ Wilkes [1988], 9.

²¹⁴ Mach [1960], 34.

²¹⁵ Irvine [1993], 142.

Understanding the grounding relationship with respect to thought experiments in many different fields shows how certain features of the theoretical context can make grounding difficult. In ethics, for example, the theoretical context is

1. reliant on a number of disparate background theories and composed mainly of common-sense beliefs, as is the theoretical context in the philosophy of personal identity, and
2. not well-defined, in contrast to the theoretical context of physics.

5.4 Thought Experiments and Computer Simulations.

In §2.5.5, I distinguished between simulation demonstrations and simulation experiments. Simulation demonstrations are tools used in order to demonstrate a theory. An example of a simulation demonstration is a flight simulator. Simulation experiments either test the explanatory potential of a theory or causal links between one state of affairs and another. I am interested in simulation experiments, rather than simulation demonstrations, just as I am interested in thought experiments, rather than thought scenarios which illustrate a theory.²¹⁶

Recall the simulation of flocking described in §2.5.5. This simulation tests the explanatory potential of a theory about flocking. When a simulation or a thought experiment tests the explanatory potential of a theory, we need to test whether the explanation is grounded, for the experimental set-up simulates an instantiation of the events or things which are explained. The explanation is grounded to the extent that it does not posit any actions of birds that are inconsistent with our theories of bird behaviour — in other words, to the extent that it replicates the type of flocking behaviour we have previously observed and recorded. When we say an explanation is grounded, we

²¹⁶ See §2.6.

are saying that it is both consistent with other theories in its theoretical context, and that it does not postulate implausible propositions with respect to those theories. The flocking rules (do not bump into one another, keep up with others, and do not stray too far away from one another²¹⁷) indicate an explanation that meets these conditions.

Note that we do not have grounding worries about the situations or facts that are explained; it is enough that the simulation convincingly replicates the phenomenon under study. Yet, as we will soon see, there are concerns about the simplifying assumptions used to represent complex situations in a simulation.

Now, consider a complex computer simulation, one designed to examine the question, "Is it rational to be moral?" This simulation does not test the explanatory potential of a theory, rather it tests causal links between one state of affairs and another. We know that cooperative societies function better than non-cooperative ones and, at the same time, we know that individuals in cooperative societies can take advantage of others' cooperative behaviour. Cooperative behaviour sometimes requires constraint in the form of acting against one's immediate interests. However, the unconstrained agent can benefit from the cooperative agent's constraint. Danielson uses computer simulations to examine whether it is always rational to be an unconstrained maximizer. That is, Danielson is looking for a rational justification of morality by broadening our conception of rationality. His hypothesis is that other behaviours besides purely constrained moral behaviour can produce collective situations that yield benefits for their participants. There may be ways for people to act in their best interests, in a rational way, that still results in a cooperative society.²¹⁸

On the basis of his simulations, Danielson concludes that "there are moral agents which are rational in the following sense: they successfully solve social problems that

²¹⁷ Kelly [1994], 11.

²¹⁸ Danielson [1992], 3-6.

amoral agents cannot solve."²¹⁹ He establishes this conclusion by setting up computer worlds in which players, or agents, interact by partaking in games that characterize social problems. The agents have two choices, they may either cooperate or not. They act from rational self-interest, in a very narrow sense. Danielson takes some time to explain how his games are analogous to certain social situations. For example, many social problems share the structure of a game such as the Prisoner's Dilemma, one of the main games used in Danielson's simulations. By building players that act in the manner prescribed by various theories and testing these agents in artificial worlds, Danielson is able to examine the complexities that arise from these interactions.

The Prisoner's Dilemma (PD) is a two-person game in which each player chooses either to "cooperate" or "defect". The payoff matrix for their choices is as follows:

		cooperate		defect	
		Player 2		Player 2	
cooperate	Player 1		2		3
		2		0	
defect	Player 1		0		1
		3		1	

In this game, no matter what the other player does, that player's opponent would maximize his or her pay-off by defection. But, if mutual cooperation could be attained the result would be better for both. In a simultaneous game mutual cooperation is hard to achieve. In the extended Prisoner's Dilemma, one player moves first and the other

²¹⁹ Danielson [1992], 4.

responds to the move. In this case, the second player always benefits more by defecting, unless the players consult one-another beforehand and each promises to cooperate if the other does. But, the promise-making and promise-keeping aspects of this solution raise difficulties. How does the first player know whether the second will keep the promise? Moreover, why should the second keep the promise after the first has already moved, and the second would do better to defect?

In the iterated Prisoner's Dilemma (IPD), players play against one another repeatedly. Over the course of the interaction a pattern of mutual cooperation may develop. Computer tournaments of IPDs, in which a group of players play one another in a round-robin tournament, show that simple cooperators are taken advantage of, and simple maximizers do poorly. A strategy such as tit-for-tat, which returns cooperation for cooperation, and defection for defection, proves to be fairly robust.²²⁰

A further twist on the IPD simulations is added with ecological and evolutionary tournaments. In the ecological variety, agents who do well in one round of the IPD are more likely to be selected, in numbers proportional to their success, to play in the next round. As the process repeats, the changes in proportion of players' strategies are observed over generations. In this model, the environment of players' strategies changes with each generation. A strategy that does well in the starting environment may die out as the proportion of other strategies changes over generations.

In the evolutionary tournament, players consist of "genes" which contain strategy bits. Reproduction is again determined by a fitness measure. That is, those agents that do best are more likely to be bred into the next generation. In this case, the breeding consists of both crossover and mutation of the genetic strings that make up each of the players.

²²⁰ Axelrod [1984], 27f.

From the brief descriptions given above we see that there is a great variety of Prisoner's Dilemma simulations to examine: there are one-shot games, extended games, iterated games, ecological games and evolutionary games.²²¹

PD simulations challenge our basic ideas about rationality. On the narrow conception of rationality, the rational thing to do is that which results in the best pay-off. On this conception, the PD player is always rational to defect, because this brings the best outcome no matter what the other does. But, a tension arises. The PD player would do even better if it could achieve a large percentage of mutual cooperation in its individual encounters, and if it could add to the probability of an overall environment of mutual cooperation. Indeed, in iterated versions of the game we observe that it is not the always-defect player who does the best.

PD simulations also challenge our basic ideas about cooperation. The always-cooperate player is a sucker, of whom advantage is taken. In the ecological and evolutionary tournaments, the always-cooperate strategy creates an environment for the always-defect strategy to flourish. It is only when always-cooperate dies out that strategies less nasty than always-defect can gain ground. From these simulations we see that not only is the always-cooperate strategy bad for the individual agent, it also creates an environment which is detrimental to the development of cooperation for the overall population. These studies of PDs show that we need a more discerning cooperator, one who can respond to others' actions in a way that promotes cooperation in the overall population.

In his simulations, Danielson finds that although moral players successfully solve social problems that amoral players cannot solve, "rationally successful agents are not as moral as we might like, nor are they completely successful in ridding their world of

²²¹ For further explanation of these varieties of Prisoner's Dilemma, see Danielson [1995a].

amoral predators."²²² Danielson's findings are not a refutation of other theories of rationality and morality. Rather, they present an alternative way of modelling our available data and, hence, a challenge to other theories.

Danielson's simulation takes up where Gyges Ring leaves off. Gyges Ring leaves one with the question, "If people act unjustly and maximize self-interest when they can escape the consequences of their behaviour, how is it possible to have a just society and why is it rational for the individual to act justly?"²²³ Danielson is successfully able to model situations in which self-interest and moral behaviour coincide.

Of his simulation, Danielson writes,

Obviously this is a world that is barren in certain respects; in others it is praeternaturally rich. I am not trying to model a minimal set of assumptions. Rather, the idea is to set a particular problem in a maximally fruitful problem-solving environment. H*Land is a world that poses a challenge common to Hobbes and Hume in a way that allows us to use powerful tools we owe to the tradition in computer science. (Incidentally, these are the sort of logical tools of which Hobbes and Leibniz dreamed.)²²⁴

Here we get insight into the difference between computer simulations and thought experiments. Two differences, in particular, are worth noting:

1. Hobbes and Leibniz dreamed of this kind of tool because it allows us to model whole worlds. With it, we have access to a "maximally fruitful problem-solving environment." In computer simulations, we can test complicated situations by tracking more variations of variables than we can in thought. For example, we can see how various groups of agents employing various sets of strategies interact. Moreover, we can employ iterated simulations, building each simulation on the results of previous simulations.

²²² Danielson [1992], 4.

²²³ See §2.1 for a discussion of Gyges' Ring.

²²⁴ Danielson [1992], 5.

2. The computer simulation is a kind of a model. We take an hypothesis — that certain behaviours will result in certain outcomes — and then we use the results to suggest something about our behaviour in our world. As Danielson writes,

[m]y artificial game world models important real moral problems.
Therefore players successful in my world may teach us something about
how to deal with our real world.²²⁵

The model suggests one way of operating in a situation. Thought experiments in science and philosophy test aspects of theories. In applied science we often use models to test designs. Computer simulations such as Danielson's are much like models. Rather than using the simulation to test theories, we use it to test an explanation (as in the flocking simulation) or causal links between one state of affairs and another (as in Danielson's rational agents and the resulting social structure). At the same time, simulations have repercussions for theory. For example, Danielson's simulation has repercussions for the theory of rational choice which "by defining rationality as unconstrained choice, makes morality irrational by definition."²²⁶ His simulations demonstrate that rationality and morality are not exclusive behaviours.

Since the question of this thesis concerns grounding, we need to consider how it is that simulations are grounded. Simulations, on my account, test the explanatory potential of a theory about the world, or causal links between one state of affairs and another. Consider the grounding conditions for each in turn:

1. A simulation that test the explanatory potential of a theory is grounded when the theory tested is consistent with other theories in the theoretical context and the theory does not postulate implausible propositions with respect to those theories. The simulation needs to show how it is that certain facts about the world are explained by the theory by showing a mechanism that is consistent with both the facts and related theories

²²⁵ Danielson [1992], 6.

²²⁶ Danielson [1992], 3.

which link the two. The mechanism takes the theoretical propositions and instantiates them in a concrete way in order to show how the theory has explanatory potential.

When a simulation tests the explanatory potential of a theory, we need to compare that explanation with other possible explanations in order to determine which explanation is the most plausible. This weighing of competing explanations is a separate process from ascertaining grounding. Plausible accounts of explanation are given by both Kitcher and Gardenförs. For Kitcher, explanation is unification; a good explanation is one which includes the most phenomena in one's belief set with the fewest and most stringent patterns of argument.²²⁷ For Gardenförs, an explanation functions as a surprise reducer. A good explanation is one in which explaining a fact E in a given state of belief answers the question, "If I did not know that E, what acceptable sentences T or C could be used to raise the probability of E (diminish the surprise of E) without begging the question?"²²⁸ The weighing of alternate explanations is a further task to be undertaken after ascertaining that the simulation has given an explanation that is grounded in the theoretical context. This grounding is achieved when the explanation is both consistent and plausible with respect to the theoretical context.

2. A simulation that tests the causal links between one state of affairs and another is grounded when the modeled states of affairs are consistent and plausible with respect to their theoretical contexts. In these simulations simplifying assumptions play a large role. In Danielson's simulations we see that he proposes an hypothesis about rationality and morality and a desired result, a cooperative society. In this case, an hypothesis about the link between one state of affairs and another is tested.

Simulations involve simplification. In order to make a good simulation one needs both an accurate model of the system and accurate input parameters. Consider a

²²⁷ Kitcher [1981].

²²⁸ Gardenförs [1988]. 177-178.

simulation called "SimHealth" which gives the user an opportunity to design a national health care system. The goal of the program is to make complex public policy accessible to the general public.

The simulation begins with a short scenario in which the simulation user is to imagine being in a serious car accident and being disappointed with the ensuing care given by the health system. This experience leads the user to decide to run for office in order to improve the health care system.

Before running for office, the user is asked to define his or her values. The values are described as follows:

- Liberty: freedom to live and do as you please.
- Equality: treating everyone the same and providing everyone with the same opportunities.
- Community: taking responsibility for the quality of life within our communities.
- Efficiency: maximizing productivity and achieving the greatest results with the least cost and waste.²²⁹

The user is then given four conditions, each of which describes one of the four crisis points in health care: cost, coverage, choice and technology. With the aid of four opinionated statements which give a solution to the problem from the standpoint of each of the four values of liberty, equality, community and efficiency, the user is asked to set his or her values.

After selecting these values, the user is "elected". The goal of the game is to create a health care system that stays consistent with the user's chosen values. Elections every four years supply this feedback. Other feedback is available in a variety of ways. Here are a few examples:

²²⁹ SimHealth [1993], 10.

1. For values — a chart plots one's originally stated values against those that one's policies reflect.
2. For values and budget — poll readings are available.
3. For budget — one can consult records of the public accounts and see figures on items such as what percentage of household income is going to taxes.
4. For the state of society — the visual depiction of downtown shows the state of various parts of society. For example, if taxes are too high, because health care costs are higher than the economy can afford, the building depicting business appears run-down and city hall swells.

To illustrate how the simulation works, it is instructive to look at an example. Suppose that the user leans strongly toward equality in his or her chosen values. To be consistent with this choice, universal insurance coverage must be set up. This move may then lead to rising deficits and crumbling schools. In this way, SimHealth improves the user's understanding of the constraints on implementing health care policies in a system with limited resources.

An important feature of SimHealth involves the user's ability to control the underlying assumptions used in the simulations. The user can run the same health care proposals against both friendly and critical assumptions. A set of policies that work using the friendly assumptions may need to be fine-tuned to do so under the critical assumptions. Here is a specific example of how this works. Under the friendly assumptions, the decision to increase by 10 percent the number of acute care procedures covered by insurance would increase the demand for hospital beds by 2 percent; under the critical assumptions, the increased coverage would generate a 5 percent increase in

demand for beds.²³⁰ The increase in demand for hospital beds may make the chosen set of policies unviable.

The goal of SimHealth is to encourage people's understanding of the complexity of health care issues through a user-friendly medium.²³¹ While SimHealth cannot tell us what health care system to implement, it can teach us how to evaluate the complex health care system in a more fine-grained fashion.

Danielson's simulation and Simhealth are different from the simulation of flocking and the thought experiments of Darwin. In these cases we are not testing the explanatory potential of a theory. In Danielson's case an hypothesis is proposed about the relation between two states of affairs. One state is modeled and the results noted. The results are put forward as evidence for the proposed hypothesis. In the case of Simhealth various simulations of a health care plan are modeled and the results noted. Each model is assessed partly on its ability to meet the needs of the proposed users. Simulations allow the user to explore a wide range of possibilities about the likely behaviour and results of some state of affairs. In order to make a good simulation one needs both an accurate model of the system and accurate input parameters. The simulation program computes how a system will behave for each choice of parameters.

When we decide to study a particular system, we are isolating a part of the world from its environment. What we include in our model of a system and its environment depends on what we want to know about the system. That is, simulations model complex systems by employing simplifying assumptions which are chosen to best represent the aspects of the system under study. It is important to develop accurate simplifying assumptions. I noted earlier that SimHealth is unusual in that some of the assumptions

²³⁰ Passell [1993], A16.

²³¹ SimHealth [1993], 5.

are under the user's control. The control over friendly and critical assumptions gives the user a sense of how it is that the results of simulations rely on these assumptions.

Any simulation — no matter how complicated it may be — has a limited range of applicability. Great care must be taken when using simulations to predict future events. For example, a simulation commissioned by the Club of Rome — *The Limits to Growth* — has been used to predict the state of the world 50 years from when it was produced.²³² The use of this simulation as a predictive tool is contested. The main problem with this simulation is that its model of the system is too rough, and the data and parameters are insufficient. The model is too rough because it models the world as a homogenous system, in that variables such as population and land resources refer to totals for the entire world.²³³ The data are insufficient partly because the parameters reflecting the ecological and economic factors affecting the world are not easy to ascertain. Further, some factors that can have a profound effect on the state of the world are not included in the model. For example, the political system in India has not been able to bring about effective birth-control in Indian villages, while the approach taken by the Chinese, whatever its moral value, has been effective. The difference in effectiveness could be due to a multiplicity of factors involving cultural, political, economic and religious features of these two societies. The interplay of factors such as these is difficult to capture in a world model.²³⁴

Simulations allow the user both to observe and to understand how systems work. In simulations, the effects of changes are observed on whole systems, and so the domain of enquiry is broad. Thought experiments, by contrast, usually aim to examine only a particular, narrowly-defined question. While thought experiments set up a fixed relation

²³² Arbib [1977], 287.

²³³ Arbib [1977], 293. A later study, again commissioned by the Club of Rome — *Mankind at the Turning Point* — does not exhibit this error. See Arbib [1977], 293-297.

²³⁴ Arbib [1977], 289.

between the independent and dependent variables that impinge on the question being examined, simulations can allow variables to fluctuate simultaneously. That is, thought experiments work in a focussed way on a particular question through variable correlation. Simulations work more broadly; they are designed to track multiple variables as they interact in a system.

In what follows, I explore the limits of applicability of SimHealth through examining the possible ways in which users may be led astray or constrained through a lack of understanding of the advantages and disadvantages of simulations.

5.4.1 Limits of Simulations

SimHealth is a rich simulation in that it models an operational society similar to our own. The user is given a state of affairs and asked to make health care policy work. However, the SimHealth society is not the only possible social system, and it is not likely to be the system an individual designing a society in another way might create. It is important for the user to realize that a simulation even as rich as SimHealth is still conservative in that it constrains one's imagination about the system. A simulation is a model of a part of the world. The simulation presents the user with a particular model and allows manipulation within that model. Simulations do not encourage users to question the given model, or to imagine alternate models.

Simulations must employ simplifying assumptions. SimHealth focuses on the declared values of the player and the economics of the social system. The user needs to understand these foci while at the same time keeping in mind further values that may be important to him or her. For example, some users may, upon reflection, be concerned about the interpersonal attitudes of health care practitioners. In their preferred health care system, the same practitioners and technology may be used, at the same cost, and the same values of liberty, equality, community and efficiency may be held, and yet the practitioners might relate to their patients, or clients, in a radically different manner. This

change may involve changes in medical training, or in society as a whole. When we use SimHealth we need to be aware that the values included are only a few of the broad array we might consider, and that we might think of the given values in different ways from those presented by the simulation.

At the same time, the SimHealth model may be too homogeneous to allow the user to focus on some important issues. I noted this problem in relation to the Club of Rome simulation. The user may wish to focus on the health care system for a particular locale, age group, gender, or religious group. The homogeneous treatment of the health care system in SimHealth leads to concerns with, for example, data about the number of hospital beds, and the levels of doctors' fees which can best be addressed on a global scale, rather than concerns with interpersonal and intrapersonal values which can best be addressed on a face-to face scale.

The complexity of computer systems may also lead one to think that with the right juggling of factual constraints, a fair and workable system can be devised. However, consider the role of values in SimHealth. We have seen that SimHealth incorporates each user's values, and yet there may be important values left out. Some such values are those which are not best considered globally and so necessarily could not be covered by a simulation with the scope of SimHealth. Other such values are those that users may hold central but fail to recall when faced with the richness of the model. Still other possible values are those that the user might imagine, were he or she given a different working process.

We have examined a number of types of simulations. The flocking simulation tests the explanatory potential of a theory about bird flocking. The model of the birds according the parameters indicated by the flocking explanation is grounded because it is consistent with what we know about bird behaviour and it accurately mimics flocking behaviour. Danielson's simulations show that there are moral agents which are rational in

that they can solve social problems which amoral agents cannot. Danielson's simulations are designed to test causal links between certain behaviours and certain results. The SimHealth simulation is designed to allow people to test certain policies regarding health care against the results.

While most thought experiments test specific propositions related to a primary theory, some thought experiments, like some simulations, test the explanatory potential of a theory. Other simulations test causal links between one state of affairs and another. Both explanatory and causal simulations are grounded if what they model according to the explanation or hypothesis under examination is consistent with other theories in the theoretical context and does not posit implausible propositions with respect to those theories. However, simulations also need to be evaluated by considering:

1. competing explanations or causal links. There may be other explanations of, or causal links to the result.
2. simplifying assumptions, resulting in overly homogenous models, overly barren models or false models.
3. rich presentations of simulations which may cause one to constrain one's comprehension of the situation. For example, in presenting underlying assumptions SimHealth may encourage people to leave out other important assumptions.

5.4.2 Simulations and Thought Experiments Compared

Simulations can be either simulation demonstrations or simulation experiments.

Simulations demonstrations do not test the explanatory potential of a theory, or test a causal link between one state of affairs and another as simulation experiments do.

Instead, simulation demonstrations demonstrate a particular theory and its outcomes. An example is a flight simulator. The distinction I draw between simulation demonstrations and simulation experiments is similar to the one I draw between thought scenarios, such

as Maxwell's demon, and thought experiments.²³⁵ The former illustrate a theory, the latter test propositions of a theory, or the explanatory potential of a theory.

I have been concerned, in this section, with simulation experiments. I have been calling them "computer simulations". How do they compare with thought experiments? My answer is that the difference between computer simulations and thought experiments stems from a difference between computers and human brains. The calculating capacity of the computer provides a valuable tool for enhancing the scope of thought experiments. As a consequence of this enhancement, computer simulations can serve functions which thought experiments cannot. Consider the following points:

1. A computer simulation can be used to model complex worlds that cannot be modeled in thought.
2. A computer simulation can be used to vary variables and modify assumptions more quickly and more finely than is possible using a thought experiment.
3. A computer simulation can be used to vary variables simultaneously.

The first two points show how computer simulations can be more powerful than thought experiments. I have noted earlier the possible problems inherited with the ability to model complex worlds and assumptions. The first is that the models and assumptions may not be accurate, and the second is that the models and the assumptions may cause one to overlook other important aspects of the world which is being modeled. Thought experiments must also take care on this count, but here the problem is more manageable since the analogies used in thought experiments are more simple than the models used in computer simulations.²³⁶

²³⁵ See §2.6.

²³⁶ I examined the benefits and problems of arguments by analogy in §2.5.2.

The third point highlights a difference between computer simulations and thought experiment which affects the function of computer simulations. There are thought experiments that test the explanatory potential of a theory in the same way that certain types of computer simulations do. For example, Darwin's thought experiments test the explanatory potential of his theory in the same way that a computer simulation about flocking tests the explanatory potential of a certain theory of flocking. However, there are kinds of computer simulations which serve a function that has no parallel in thought experiments. These are the simulations which suggest causal links between a certain state of affairs and another. Danielson's simulations and the SimHealth simulation suggest ways to proceed in order to achieve certain ends. The computer simulation is not necessarily used to vary specific variables, for it may change groups of variables simultaneously, with a the purpose of modelling a situation that produces certain desired results.

In "New Computers, New Thoughts",²³⁷ James Bailey claims that the working processes of computers are changing the way we think. He holds that, to date, it seems that the human mind works best sequentially, doing one task at a time, and that this fact has led people to impose the notion of sequentiality on a nonsequential world.²³⁸ Various developments in computer modeling have helped us to experience parallel, rather than sequential processes. Computer graphics, lattice gas algorithms and genetic-algorithm computation all model behaviour of systems non-sequentially.²³⁹

Both Danielson's work and SimHealth are examples of the way in which computer developments can help us to "see" how complex aspects of systems interact and so think in new ways about systems. They are able to model simultaneous interactions

²³⁷ Bailey [1992], 232.

²³⁸ Bailey [1992], 232.

²³⁹ Bailey [1992], 234.

with a complexity that the human brain alone appears unable to manage. It is in this ability to model "maximally fruitful problem-solving environments"²⁴⁰ that computer simulations differ most significantly from thought experiments.

²⁴⁰ Danielson [1992], 5.

6. Grounding Responses

The response or reaction to a thought experiment can be distinguished from the result of the thought experiment. In the argument form of the thought experiment the responses are included either as premisses or subconclusions, leading to the result, or conclusion. For example, recalling again the argument contained in the ill-violinist thought experiment (§2.5.1), we see that the reaction to the thought experiment is given at step 12, and forms part of the argument leading to the result at step 15.

The criteria for successful thought experiments, like those for successful physical experiments, ought to be independent of the results.²⁴¹ On my account a thought experiment is successful if it employs strong reasoning and is grounded. Features of the result of the thought experiment do not indicate the success of the thought experiment. That is, we may have a successful thought experiment, in that the thought experiment itself and the response to the thought experiment are grounded, and yet there may be no useful result achieved. Or it might be that a thought experiment that is not grounded accidentally yields correct results.

Consider the Galileo thought experiment given in §2.3. We have seen that while the characters in the dialogue clearly refer to the objects falling in a medium, modern readers are able to dismiss the medium through which the objects fall and so reach the conclusion applying to bodies falling through a vacuum.²⁴² When we examine the experiment we see that it yields a correct result for objects falling in a vacuum and yet it overlooks the important issue of the medium, which is the situation given in the actual context. The thought experiment achieves a correct result, but when considered in the theoretical context relevant to the dialogue, it is not grounded.

²⁴¹ Unger acknowledges this as a potential problem for his methodology. See Unger [1990], 14.

²⁴² Norton [1996], 343-345.

In this chapter, I show that responses to thought experiments must be examined closely. Responses to thought experiments are often called intuitive responses. Some mistrust these responses, others are willing to take them seriously. We can see evidence of these two contrasting positions in examining the views of Thomson and Unger.

On Thomson's account we employ thought experiments in ethics because:²⁴³

1. We do not know what accepting any moral principle entails for our behaviour until we examine how it plays out in a concrete situation, and
2. Our views about stories and examples give us data for moral theorizing; we use stories to see if our moral principles or theories explain the data or predict future data.

When Thomson claims that our responses give us data for moral theorizing, I see it this way. Our responses give us the data about our beliefs about our values. These data are then considered and used to further modify our beliefs arising from our frameworks of values.

Unger outlines two different approaches to thought experiments:

1. The Preservationist approach is that

[a]t least at first glance, our moral responses to particular cases appear to reflect accurately our deepest moral commitments, or our *Basic Moral Values*, from which the intuitive reactions primarily derive; with all these case-specific responses, or almost all, the preservationist seeks to *preserve* these appearances. So, on this view it's only by treating all these various responses as valuable data that we'll learn much of the true nature of these Values, and, a bit less directly, the nature of morality itself.²⁴⁴

2. The Liberationist holds that

folks' intuitive moral responses to many specific cases derive from sources far removed from our Values and, so, they fail to reflect the Values, often even pointing in the opposite direction. So, even as the Preservationist seeks (almost) always to *preserve* the appearances promoted by these

²⁴³ Thomson [1986], paraphrased from 257.

²⁴⁴ Unger [1996], 11.

responses, the Liberationist seeks often to *liberate* us from such appearances.²⁴⁵

Unger describes two different attitudes to beliefs about values here. In the first, beliefs are accepted at face value. In the second, the beliefs are subjected to rigorous questioning. On my account, we try to ground our responses through careful consideration of their relationship to the theoretical context in order to avoid the shallow acceptance of first-blush responses.

Here, we will see that the quality of our responses depends partly on our knowledge of the theoretical context to which they relate, partly on the way that the thought experiment is presented and partly on our awareness of the role of subjective filters.

Hon claims that there are two sources of errors in observational reports in physical experiments. First, observational reports may introduce misinterpretations of what has been correctly observed. Alternatively, they may reflect limitations of the senses.²⁴⁶ An example of the first type is apparent in a situation in which one fails to account for the external circumstances which influence an observation. For example, in making a pointer reading, the gap between the pointer and the scale must be considered in order to account for the phenomenon of parallax.²⁴⁷ In thought experiments, the experimenter must understand how his or her response is affected by the presentation of the thought experiment and its context in relation to other thought experiments.²⁴⁸ A mistake of the second type is evident in a physical experiment when an experimenter makes an error due to limitations of the senses. Clear examples are errors due to colour blindness or slow

²⁴⁵ Unger [1996], 11-12.

²⁴⁶ Hon [1989], 489.

²⁴⁷ Hon [1989], 489.

²⁴⁸ See Chapter 7.

reaction time.²⁴⁹ Other examples of errors due to the subjectivity of the senses are found in situations in which anticipation and hallucination influence observations.²⁵⁰ In thought experiments the role of subjectivity is more pronounced than it is in physical experiments, since responses are formed in relation to individual reasoning and psychological processes. Furthermore, we shall see that although inter-subjective agreement about responses may help to minimize distortions, it can not guarantee a lack of distortion.

The following example of the role of the limitation of the senses and subjective anticipation in the interpretation of observations is given by Hon.²⁵¹ The canals of Mars were first labelled such by Schiaparelli of Milan. Later, Lowell and others proposed the theory that there exists a highly intelligent civilization on Mars. Lowell hypothesized that the canals were built to transport water from the polar ice caps to the equatorial region of the desert-like planet. Lowell admitted that the canals were difficult to see and that observation of the canals required a trained and acute eye operating under the best conditions. The Mariner pictures made it clear that the canals were a fiction. Hon writes,

It appears that poor resolution initiated a condition which facilitated a visual synthesis of discontinuous elements; the interpretation of these elements as canals was supported in turn by the belief that there is a Martian community. Thus, the observation of the canals was in error due to physical-physiological limitation as well as psychological expectation.²⁵²

The account of the observational error here provides a way to think about errors in our responses to thought experiments. Our responses are constrained by our ability to

²⁴⁹ Hon [1989], 491.

²⁵⁰ Hon [1989], 491.

²⁵¹ Hon [1989], 492f.

²⁵² Hon [1989], 493.

reason and be conscious of our reasoning process. Our responses are also influenced by psychological expectation and subjective distortions.²⁵³

Hon raises another interesting point about observation. Here, he writes of the phenomena of observing N rays, which turned out to be resistant to attempts to duplicate:

By definition, an observation is a subjective process; the problem lay rather in the nature of the observation: it was simply unamenable to objective, or rather inter-subjective criteria such as pointer reading of measurement or the length of a spark.²⁵⁴

In thought experiments we may achieve inter-subjective agreement, however, it is not intersubjective agreement about a sense perception that is measured, but rather about a response. In the examples of the canals of Mars and N-rays, we see that even physical observations can be influenced by intersubjective agreement borne of strong ideas and expectations. Agreement about ideas in relation to theory is often more susceptible to group influence than agreement about physical observations. Furthermore, thought experiments in ethics are conducted against the framework of values, a framework that is partly formed by inter-subjective influences. So we are trying to monitor inter-subjective agreement about a framework for which there is already some inter-subjective agreement. Some may feel that this will get us nowhere, since we cannot get outside of the system. However, this problem is not unique to thought experiments concerning ethics. We cannot get outside of scientific paradigms either. In both cases, we do the best we can.

One criticism of taking responses to thought experiments seriously is that they are only the responses of a person of a particular group and so they reflect the theoretical and pre-theoretical biases of that person and group. Sorensen summarizes this kind of criticism of thought experiments in the following passage:

²⁵³ We will see examples of these kinds of distortions in my discussion of personal identity and ethics thought experiments in this chapter.

²⁵⁴ Hon [1989], 494.

Thought experiments are alleged to articulate the intuitions of an unrepresentative group such as Oxford dons, upper-class white males, or those who fund theorists. The theory-ladenness of imagination is used to smirk away thought experiments as epiphenomenal, circular, and self-deceptive. The constancy of their verdicts is explained as a kind of collective wish fulfillment, as components of a social fantasy system that is driven by desire rather than truth.²⁵⁵

Our responses can not help but be subjective and somewhat representative of our group. The thought experiment's scope of application must be considered in light of this fact. In other words, we are using thought experiments to explore and clarify responses about a particular subject as it is understood in a particular group or by a particular person. In thought experiments regarding values, such as those concerning ethics, this is particularly evident, but from the contextualist point of view the best that can be done in thinking about ethics is to clarify one's approach in relation to one's context, which means narrowing one's focus to the particular people involved and their framework of values in relation to the particular situation under consideration. At the same time, comparing our responses to a thought experiment against those of others from outside our group can bring a new perspective to the thought experiment. If inconsistencies are noted in the responses, the result can be a modification in the framework of values, a modification which may not have occurred if the comparison were not made.

In this chapter, I suggest that responses to thought experiments, if they are to be useful, must be grounded. A successful thought experiment process is one in which the experimenter employs considered responses rather than first-blush intuitions.

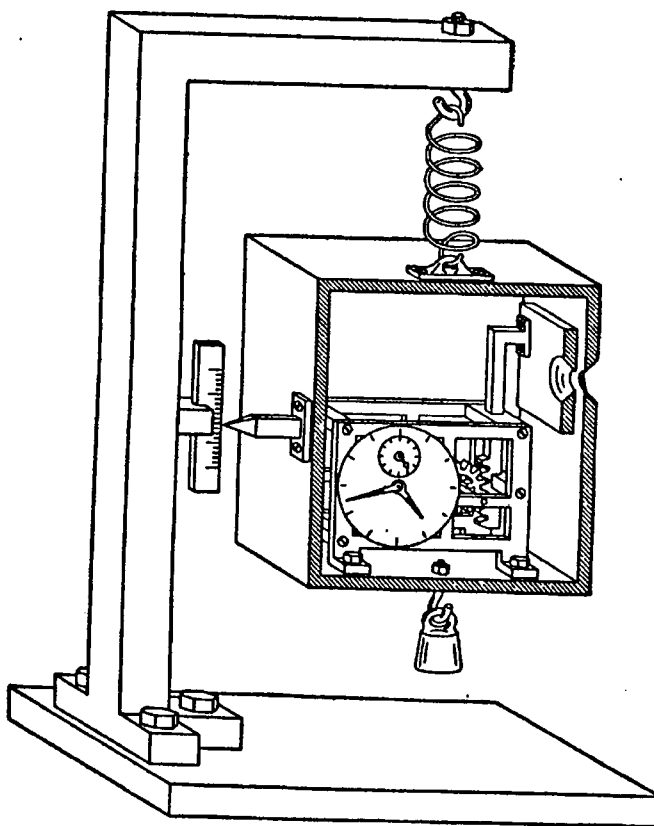
6.1 Responses and the Theoretical Context

In this section, I suggest that responses vary depending on two factors relating to the theoretical context. The first has to do with the individual's knowledge of the theoretical

²⁵⁵ Sorensen [1992a], 235.

context, the other with the state of development of the theoretical context itself. I now consider each factor in turn.

Responses can vary in relation to the individual's knowledge of the theoretical context. For example, in conducting a scientific thought experiment, one needs a considerable amount of scientific knowledge. A response of a person without this knowledge may be suspect. Consider the following thought experiment. In 1930 Einstein devised the "clock in a box" thought experiment to show a "contradiction to the reciprocal indeterminacy of time and energy quantities in quantum mechanics."²⁵⁶ The apparatus of the thought experiment is shown in the following diagram:²⁵⁷



²⁵⁶ Bohr [1959], 226.

²⁵⁷ Bohr [1959], 227.

The clock controls a shutter covering a hole in the side of the box. The box is suspended on a spring balance with a pointer protruding from its side. This pointer measures the box's position relative to a scale attached to the supporting structure of the spring balance. The box contains a gas of photons and the clock is set to open the shutter at the precise moment when a single photon could be released through the hole. By the formula $E=mc^2$ one could, by weighing the box before and after the release of the photon, obtain the energy of the photon. Einstein suggested that the ability of this imagined apparatus to determine the precise moment of the release of the particle and measure the energy of the photon pointed to a "contradiction to the reciprocal indeterminacy of time and energy quantities in quantum mechanics."²⁵⁸

Here is part of Bohr's discussion of the clock in a box:

The box ... is suspended in a spring balance and is furnished with a pointer to read its position on a scale fixed to the balance support. The weighing of the box may thus be performed with given accuracy Δm by adjusting the balance to its zero position by means of suitable loads. The essential point is now that any determination of this position with a given accuracy Δq will involve a minimum latitude Δp in the control of the momentum of the box connected with Δq by the relation (3).²⁵⁹ This latitude must obviously again be smaller than the total impulse which, during the whole interval T of the balancing procedure, can be given by the gravitational field to a body with mass Δm , or

$$\Delta p \approx \frac{h}{\Delta q} < T \cdot g \cdot \Delta m, \quad (6)$$

where g is the gravity constant. The greater the accuracy of the reading q of the pointer, the longer must, consequently, be the balancing interval T , if a given accuracy Δm of the weighing of the box with its content shall be obtained.

Now, according to general relativity theory, a clock, when displaced in the direction of the gravitational force by an amount of Δq , will change

²⁵⁸ Bohr [1959], 226.

²⁵⁹ Relation (3) is $\Delta q \cdot \Delta p \approx h$ where h is Planck's constant. This is Heisenberg's indeterminacy principle. See Bohr [1959], 209.

its rate in such a way that its reading in the course of a time interval T will differ by an amount ΔT given by the relation

$$\frac{\Delta T}{T} = \frac{I}{c^2} g \Delta q. \quad (7)$$

By comparing (6) and (7) we see, therefore, that after the weighing procedure there will in our knowledge of the adjustment of the clock be a latitude

$$\Delta T > \frac{h}{c^2 \Delta m}$$

Together with formula (5)²⁶⁰, this relation again leads to

$$\Delta T \cdot \Delta E > h,$$

in accordance with the indeterminacy principle. Consequently, a use of the apparatus as a means of accurately measuring the energy of the photon will prevent us from controlling the moment of its escape.²⁶¹

In effect, Bohr's response to this experiment is that it would not achieve what Einstein predicted. This is because, on the release of the photon, a momentum will be transferred to the box. According to Einstein's theory of general relativity the resulting motion in the gravitational field would affect the rate of the clock. The effect on the clock is subject to the indeterminacy principle since it arises from the momentum of the box which is a result of the momentum of the photon.

It is evident that with respect to Einstein's clock in a box experiment, a person without appropriate knowledge of the background theory would be unable to respond in any useful way. This person would not understand how it is that the apparatus would be unable to measure accurately the energy of the photon and the moment of its escape.

We say that a response is grounded if it arises from and accords with the theoretical context. In thought experiments in ethics, responses arise from knowledge of

²⁶⁰ Formula (5) is $E = mc^2$. See Bohr [1959], 225.

²⁶¹ Bohr [1959], 226-228.

a framework of values, which unlike specialized scientific knowledge, is generally accessible to the majority of people.

Responses to thought experiments can vary in relation to the state of development of the theoretical context. In thought experiments in which the theoretical context is well-defined our responses tend to be fairly constant. In those in which it is not they can be extremely varied. We can see the instability of intuitions in the following examination of thought experiments about personal identity.

Unger is concerned with the theoretical grounding of personal identity thought experiments. He holds that personal identity thought experiments must be grounded in the following two ways:²⁶² 1) our intuitions from thought experiments must be weighed against plausible philosophical hypotheses and 2) the thought experiment itself must conform to the basic scientific principles which are central to our beliefs.

Condition 1) rules out certain responses to thought experiments. Here Unger is not concerned with the grounding of the thought experiment itself, but with the grounding of our responses to the thought experiments. Our responses, he claims, need to be checked against plausible philosophical hypotheses.

Condition 2) rules out certain types of thought experiments as ungrounded. I have stipulated criteria for grounding the thought experiment in its theoretical context in Chapter 3.

Unger claims that our reactions to his examples cannot give us access to the main facts about our existence and survival. Rather, these facts must be discovered by empirical labour. Thought experiments can guide us in exploring our deeply held beliefs

²⁶² Unger [1990], 8.

about survival,²⁶³ beliefs that have been shaped by scientific investigation.²⁶⁴ It is from these deeply held beliefs that our responses to thought experiments arise.

There are, however, problems which present themselves when we consider Unger's proposal. They are

- i. Who are the "we" who have the deeply held beliefs?
- ii. What if our deeply held beliefs are inconsistent?

Unger writes that the "we" are those colleagues and students who are both detached and astute. We must be detached in the sense that we should not be attached to any particular approach. Swinburne takes exception to the broad generality of Unger's approach. He claims that the "deeply held beliefs" that Unger ascribes to "us" are not held by many people. Swinburne, naturally, counts himself as one of those many people.²⁶⁵ From examining Unger's work it is clear that he is exploring the deeply held beliefs of those who accept a naturalistic framework for their world view.

A more serious problem with Unger's approach lies in the fact that people's responses, based on their deeply held beliefs, are often inconsistent. Within any group of people, even a group whose beliefs are based on current scientific theories, conflicting responses can arise. For example, if one person imagines the information in the brain can be stored and transferred through a device that makes a tape, and another person imagines that it is stored in a holographic device, their responses about various thought experiments involving transferring information from one brain to another may differ. Further, even if two people think that the taping device is plausible, one person's responses may be informed by visions of an analog process and the other's by those of a

²⁶³ Unger [1990], 3.

²⁶⁴ Unger [1990], 10.

²⁶⁵ Swinburne [1992], 149.

digital one. Responses will be led astray by the diversity of people's imaginings, inconsistency between people's beliefs and even inconsistencies that may arise within one person's beliefs.

White holds that the results Unger derives from his thought experiments are not as stable as he supposes since the experiments lack detailed specifications of the social contexts in which the examples are imagined to occur. A conventionalist account, such as Unger's, needs to pay special attention to social contexts, for it is within these contexts that conventions are established. Our reactions to things like fission, super freezing and the like would be different if we lived in a society where these things happened routinely. In this case, these practices would be coherently integrated into the culture.²⁶⁶ In our case, we do not have consistent or stable intuitions on these matters because we do not encounter them.

There is something right about this criticism. However, I do not think that it can be applied to Unger's thought experiments. I agree that in conventionalist accounts the social context of the thought experiment is particularly important. However, White has misconstrued Unger's social context. Unger does not want to test what our reactions to various processes would be were we in some society in which they were routine. Rather, he uses the unusual processes he describes to test the limits of our current society's concepts.

Unger's context is something like the following. Imagine that our society is just beginning to deal with cases such as super freezing and that we are working on a way to conceptualize them. We would, in this case, begin with the concepts we presently employ and see how they could accommodate the new cases. In the process, hopefully, our understanding of personal identity, *as it matters to us now*, would be deepened.

²⁶⁶ White [1992], 158.

Unger is not asking us to imagine we are in a social context which already routinely experiences the kinds of procedures he describes.

In sum, we have seen that Unger requires that responses to thought experiments are grounded in plausible philosophical hypotheses and that the experiments themselves are grounded in basic scientific principles that form the basis of our beliefs. While these grounding conditions seem correct, confusions about the explicit content of the deeply held beliefs can lead to problems. White's criticism of Unger gives us insight into the manner in which responses must be grounded, not only in scientific and philosophical contexts, but also in social ones. In this section, we have seen how responses to thought experiments vary in relation to an individual's knowledge of the theoretical context and also in relation to the state of development of the theoretical context.

6.2 Subjective Distortions: Personal Identity Thought Experiments

In the examples about canals on Mars, we saw how psychological factors such as expectation can affect observations. In *Human Beings*²⁶⁷ Johnston notes that there are two psychological factors that may predispose one to interpret personal identity thought experiments to favour a psychological view:

1. The psychological continuer effect: In daily life, what is important to us is our memories, experiences and intentions. Since we place a high value on the continuity of these things, and view death, in part, as the cessation of the continuity of these things, we are disposed to trace individuals in accord with psychological criteria.²⁶⁸
2. The social continuer effect: We tend to trace a person in terms of his or her social continuer. The social continuer is an aspect of the public image, or persona, of a person. The continuation of a persona is often confused with the continuation of a person,

²⁶⁷ Johnston [1987].

²⁶⁸ Paraphrased from Johnston [1987], 80-81.

particularly if this persona seems to be capable of remembering, understanding and carrying on relationships with those people in the original person's social circle.²⁶⁹

Johnston claims that the psychological and social continuer effects feed psychological accounts such as that of Parfit. He observes that while these effects trace persistence of personæ, they do not trace the persistence of persons. Persons may outlive, or be outlived by, their personæ.²⁷⁰ The effects Johnston describes do not count against psychological views, but rather against the unexamined use of the thought experiment methodology to arrive at such views.

Johnston is interested in background common-sense assumptions that affect responses to thought experiments. He thinks that these background assumptions support psychological accounts. However, I would counter with the observation that we should have a theory of personal identity that is consistent with our every day ascriptions of personal identity. Surely everyday intuitions such as the psychological and social continuer effects need to be taken into account in any developing theory of personal identity that would be both generally acceptable and applicable. The background assumptions of psychology and social practices are pertinent to the theoretical context of personal identity thought experiments.

In *Identity, Consciousness, and Value*, Unger presents an examination of the psychological factors that may influence the outcome of thought experiments in general, and personal identity thought experiments in particular.

Unger writes that a major danger to a balanced thought experiment methodology arises from contextual distortions which, in turn, may lead to psychological distortions. For example, a thought experiment told in too story-like a manner will elicit passive and

²⁶⁹ Paraphrased from Johnston [1987], 82.

²⁷⁰ Johnston [1987], 82.

non-critical audience attitudes, since most people are accustomed to going along with science fiction stories for the sake of the story.²⁷¹ Unger warns that, even in favourable non-story-like contexts, one must be aware of responses that are influenced by dominant psychological tendencies rather than interesting philosophical attitudes and beliefs. For example, Unger claims that people tend to oversimplify and overgeneralize.²⁷²

Unger presents avoidance of future great pain (AFGP) tests as a way to avoid psychological tendencies that distort responses to personal identity thought experiments. Unger claims that his AFGP tests nearly always deliver highly similar responses to the questions they pose regarding personal identity over time.²⁷³

An example of an AFGP test is this. Imagine you are going to be super frozen for three years and then thawed. Further you have the choice of 1) enduring considerable pain before the freezing and having the person who is thawed endure no pain, or, 2) facing no pain now, but knowing that the thawed person would experience a great deal of excruciating pain for a long time.

Responses to the test above would read something like this. If one chooses 1), then one assumes that one will survive super freezing. If one opts for 2), then one does not believe one will survive super freezing. Unger claims that careful consideration of how super freezing leads one to opt for 1) or 2) should give us a clearer understanding of what is needed for a theory of personal identity over time. The super freezing AFGP test can then be compared with AFGP tests that use different processes as their foci.

In Unger's AFGP tests the following conditions must be met:

²⁷¹ Unger [1990], 20.

²⁷² Unger [1990], 15.

²⁷³ Unger [1990], 27.

1) One must imagine that one will oneself undergo the process, that one is not masochistic and that one's basic attitude is one of self-concern.²⁷⁴

2) The future pain envisioned must be enormous.²⁷⁵

In condition 1), the stipulation that the respondent must consider himself or herself the subject of the experiment creates a context which engenders serious consideration of the case and leads to less casual responses.²⁷⁶

Condition 2) has two results. First, it focuses one on the question of whether one survives, rather than on related questions that may be conflated with it, such as how one might feel about the sort of person one might become after undergoing such a process and how one might feel about the sort of life one would lead in the altered state. Second, it supplies an answer to the objection that the subject of the test may be influenced by uncertainty over the outcome, or by a tendency to discount the undesirability of future pain. The enormity of the pain would help to override discount or uncertainty rates in those prone to attempting these types of calculations.

Unger's description of AFGP tests shows how he uses his awareness of possible psychological distortions to aid in developing thought experiments that will be more likely to lead to interesting philosophical results.

Unger is a physicalist. He argues that there is personal identity over time when there is continuity of core psychological capacities supported by a sufficiently continuous physical realizer.²⁷⁷ We have seen that Johnston holds that there are two common psychological distortions to personal identity thought experiments: the psychological continuer effect and the social continuer effect. The psychological tendencies Johnston

²⁷⁴ Unger [1990], 28.

²⁷⁵ Unger [1990], 29.

²⁷⁶ Unger [1990], 28.

²⁷⁷ Unger [1992a], 134.

outlines have to do with one's distinctive psychology, the continuity of which is central to psychological views.

Upon examination we see that Unger's AFGP tests negate the psychological continuer effect. The psychological continuer effect sways one's responses to reflect the tracing of distinctive psychology through changing bodies, and teletransporters and the like. The AFGP tests focus on pain, a capacity of both core and distinctive psychologies. The respondent to Unger's AFGP tests is focused on whether he or she survives, not on the continuation of his or her distinctive psychology.

Unger's AFGP tests also negate the social continuer effect. We trace people by reference to their social continuer, which is also a function of one's distinctive psychology. The AFGP test focus on pain, a capacity of both core and distinctive psychology. Again, the AFGP test focus the question on survival of core rather than distinctive psychology.

Has Unger, in constructing thought experiments that are not influenced by the social or the psychological continuer effect, suppressed these effects artificially? We would consider that he has if we held that these effects are not erroneous psychological tendencies, but rather tell us something important and true about personal identity. That is, we may consider the fact that we are prone to trace personal identity through psychological and social continuers as evidence for a particular view, rather than as problematic psychological tendencies.

Unger may agree that the psychological and social continuer effects do seem to play a part in our reactions to many personal identity thought experiments. If he does, he may argue that since his thought experiments yield responses that support the physicalist view, they have been successful in overriding these psychological tendencies, and that in overriding such tendencies, the results of these thought experiments are worth particularly serious consideration. However, should Unger choose to present this argument, one

might respond that perhaps the social and psychological continuer effects have been artificially suppressed through his careful construction of the thought experiment. The fact that the thought experiment seems to yield a particular view, in this case a physicalist one, cannot be construed as evidence that the thought experiment has been well constructed. We see that thought experiments about personal identity can be constructed against very different theoretical backgrounds (e.g., physicalist or psychological views) with only our responses which are influenced by our basic beliefs to judge between them.

Thus, the same thought experiment can be interpreted in two very different ways given the theoretical background which the experimenter assumes. In the following example we see this very clearly. Mediaeval theorists used the same thought experiment to argue for and against the possibility of a vacuum.

... vacuists claimed that a void could form within a collapsible container; just empty it, make it airtight, and force it back into its large uncollapsed shape. Vacuists admitted that this container could not be an ordinary one (such as a wine pouch) because it would leak. The force could not be the tugs of ordinary hands because they are not strong enough. So the vacuists just stipulated the container leak-proof and the forces strong enough. Plenists would use the same scenario to demonstrate the *impossibility* of a vacuum! They invited readers to picture the comical futility of trying to seal the container *and* enlarge it. If you succeed in enlarging the container, you can be sure of a leak; so you reinforce it. Now you need more enlarging force. You get it; but your success in enlarging the container shows you have a new leak. Thus, the vacuum maker is doomed to an endless spiral of ever stronger sealants and ever stronger enlarging forces.²⁷⁸

The problem here is that the thought experiment is conducted against two divergent contexts. In the one context, vacuums are possible, and so there must be containers that can hold them and forces that are able to produce them. In the other context, vacuums are impossible, and so there can not be containers to hold them and forces that are able to produce them. There needs to be some way of deciding independently of the thought experiment which context is the right one. The thought experiment cannot decide

²⁷⁸ Sorensen [1992a], 27-28.

between them. The fact that this thought experiment can be argued both ways leads to a need for development of theory that will decide between the two cases.

We have seen that both Unger and Johnston hold that we can best compensate for psychological distortions if we anticipate them and attempt to construct thought experiments that avoid undue influence from them. However, upon examining Johnson's and Unger's warnings of various psychological distortions in our responses to personal identity thought experiments, we see that these possible distortions to responses reflect the underlying disagreements in the developing theory. Not only are thought experiments difficult to ground in personal identity theory, responses to these thought experiments are difficult to ground. The reason for this is the lack of clarity in the developing personal identity theory.

6.3 Subjective Distortions: Thought Experiments in Ethics

Thought experiments in ethics operate in relation to frameworks of values. How is it that we should entertain responses arising out of such a framework? I suggest we should aim for grounded responses. Grounded responses in ethics bear some resemblance to Rawls' considered judgements, about which he writes:

... So far, though, I have not said anything about considered judgements. Now, as already suggested, they enter as those judgements in which our moral capacities are most likely to be displayed without distortion. Thus in deciding which of our judgments to take into account we may reasonably select some and exclude others. For example, we can discard those judgements made with hesitation, or in which we have little confidence. Similarly, those given when we are upset or frightened, or when we stand to gain one way or the other can be left aside. All these judgments are likely to be erroneous or to be influenced by an excessive attention to our own interests. ... the relevant judgments are those given under conditions favorable for deliberation and judgment in general.²⁷⁹

²⁷⁹ Rawls [1971], 48-49.

On Rawls' account, considered judgements must be those in which we have confidence, those which arise from a mental state in which we are not upset or frightened, and those which arise from an impartial standpoint.

Taking each in turn, it seems obvious that we cannot speak well of a judgement simply because it is made with confidence. As we know, doubts often give rise to more considered judgements. A good strategy might be to doubt moral judgements as rigorously as one can in order to obtain the most considered judgement.

The next two features of considered judgements, according to Rawls, are that they arise from a calm mental state (a non-upset, non-frightened, mental state) and an impartial standpoint. These ideas might usefully be considered in relation to Hare's remarks about thought experiments designed to argue against utilitarianism. Hare distinguishes between intuitive and critical thinking about moral problems. Intuitive thinking is the thinking that occurs when one is in an actual situation and has little time for reflection, while critical thinking occurs during calm reflection. Hare writes that

critical principles are what would be arrived at by leisured moral thought in completely adequate knowledge of the facts, as the right answer in a specific case. ... Intuitive (or, as I have elsewhere called them, *prima-facie*) principles are inculcated in moral education; but the selection of the intuitive principles for this purpose should be guided by leisured thought, resulting in critical principles for specific considered situations, the object being to have those intuitive principles whose cultivation and general acceptance will lead to actions in accord with the best critical principles in most situations that are actually encountered.²⁸⁰

On Hare's account, the commonest strategy of opponents of utilitarianism is to use thought experiments to conflate the two types of thinking so that

[i]t makes the utilitarian look like a moral monster. The anti-utilitarians have usually confined their own thought about moral reasoning ... to the intuitive level, the level of everyday moral thinking on ordinary, often stressful occasions in which information is sparse. So they find it natural to take the side of the ordinary man in a supposed fight with the utilitarian whose views lead him to say, if put at the disconcertingly unfamiliar

²⁸⁰ Hare [1989], 221.

standpoint of the archangel Gabriel, such extraordinary things about these carefully contrived examples.²⁸¹

Hare argues that the impartial standpoint can give rise to problems when applied to a specific case because it cannot account for the intuitive understanding of what, *in that specific case*, would be the right response. However, if one is given time to reflect, using one's critical capacity, one is well able to account for the intuitive response to such situations, in newly formulated critical principles. Hare describes the process this way:

It will be done by using critical thinking to consider cases, both actual and hypothetical, which crucially illustrate, and help to adjudicate, disputes between rival general principles. But, because the general principles are being selected for use in actual situations, there will have to be a careful proportioning of the weight to be put upon a particular case to the probability of its actually occurring in the lives of the people who are to use the principles. So the fantastic cases that are so beloved of anti-utilitarians will have very little employment in this kind of thinking (except as a diversion for philosophers or to illustrate purely logical points which is sometimes necessary). Fantastic unlikely cases will never be used to turn the scales as between rival general principles for practical use. The result will be a set of general principles, constantly evolving, but on the whole stable, such that their use in moral education, including self-education, and their consequent acceptance by the society at large, will lead to the nearest possible approximation to the prescriptions of archangelic thinking. They will be the set of principles with the highest acceptance-utility.²⁸²

What Hare seems to be saying is that we have intuitive responses that become a factor in forming our critical principles, which, in turn, are learned and form a basis for further intuitive responses in real-time situations.

Now, consider thought experiments in ethics. We consider whether the thought experiment and our responses to it are grounded by considering the theoretical context of the thought experiment. To do this we consider as many factors as possible in a process of wide reflective equilibrium: possible judgements and outcomes, values, situational factors, moral principles and other background factors.

²⁸¹ Hare [1989], 222.

²⁸² Hare [1989], 223-224.

As noted, grounded responses have similarities to Rawls' considered judgements. It also seems right to say that we should make our responses to thought experiments as considered as possible by applying something like Hare's critical thinking. Hare's critical thinking process resembles the wide reflective equilibrium process in which responses are to be considered against a variety of background considerations. The result is responses *for that particular case*, situated in its unique way. These particular responses are not principles to be applied in every case, but rather particular judgements that may be used when considering similar cases. The judgements cannot be called principles because they are too specific to be applied in generality. At the same time, they are principled judgements in that they embody principles which have been considered in the process leading to the specific judgement.

The more often the process of testing and reflecting on one's responses occurs, the more refined, or grounded responses become. So, in the process of conducting thought experiments concerning ethics there is potential for the development of a framework of values. The point here is that the features of mental states that Rawls and Hare claim give rise to considered judgements, can be considered as features that may influence responses to thought experiments. Judgements can be swayed by both partiality and impartiality; lack of confidence and over confidence; a non-calm mental state and an overly calm mental state. Responses to thought experiments can be swayed by the same things. Some cases need an impartial standpoint, while in some an impartial standpoint would not adequately account for unique situational aspects of the problem. Some situations are resistant to confident judgements. In fact, most difficult ethical problems are difficult because confident judgements about them cannot be attained. Most of the time a calm mental state is a pre-requisite to clear thinking while occasionally moral outrage can serve as an indication of the severity of the case and the need for a certain kind of action. The reflective equilibrium process gives a way to balance these concerns. There is no way to formulate the best way to arrive at a judgement, but a process such as that of wide

reflective equilibrium can help one to reach the balance between various considerations. Rawls' considered judgements and Hare's critical thinking indicate some of the kinds of considerations to factor into the wide reflective equilibrium process.

The criticism that might be raised here is that responses to thought experiments reflect the moral judgements of an individual rather than that of some larger group. I suggest we begin by taking the individual response to the thought experiment as a reflection of a framework of values not only of that individual, but also of the larger group to which they belong. Then we test this supposition in conversation with others about the thought experiment, in the course of which the response is seen to be consistent with that of others or not, and so with the framework of values of the group or not.

However, even if responses reflect the values of a larger group we might worry about why we should accept them. Many argue against accepting arguments from received opinion in favour of a more considered approach. The fact that thought experimenters seriously entertain responses in the process of achieving their results is not a reason to dismiss thought experiments. It is true that responses reflect the experimenters' understanding of the theoretical context of the thought experiment. For example, in ethics, responses tell us about the experimenters' beliefs arising from their framework of values and any ethical theory involved in the thought experiment. In science, responses tell us about the experimenters' understanding of scientific theory. Earlier, we saw how it is that Unger and Johnson suggest avoiding distorting influences that affect our responses to thought experiments regarding personal identity. In examining responses to thought experiments in ethics it is important to consider how we might similarly temper them.

Hare suggests a kind of test of responses that is similar in intent to Unger's AFGP tests. Hare holds that moral judgements and principles are both universal and prescriptive, and so in considering thought experiments it is necessary to consider oneself

taking each of the contrasting positions outlined in the thought experiment and seeing if one's response about what ought to be done is the same from either vantage point. He writes,

Because moral judgements and moral principles have the features of prescriptivity and universalizability, the person who adopts one is in effect prescribing universally for all situations of a certain (perhaps minutely specified) kind. These situations will include not only actual ones, but hypothetical qualitatively identical situations in which the roles of the agents are interchanged. When, therefore, I ask what I ought to do in such and such a situation, I am in effect asking for a prescription for a situation of a certain kind, on the understanding that it is to apply to *all* situations of that kind, no matter what role I myself am to play in them (for example, that of murderer or of victim).²⁸³

So the process of grounding responses would involve one in taking the roles of each individual or each position in the thought experiment and asking whether one would prescribe the same outcome in both cases.

My point here is to suggest that responses to thought experiments should be as refined as possible. The tests of Unger and Hare are suggestions for this process. There are many possible tests. It is important to question responses closely in any way that seems reasonable in order to arrive at more considered responses.

In this chapter, we have examined the role of responses to thought experiments in relation to the efficacy of thought experiments. We have seen that responses to thought experiments may be reliable, or not, in relation to the individual's knowledge of thought experiment's theoretical context (recall the non-scientist's reaction to Einstein's clock in the box), in relation to the level of development of that context (recall the difficulties in responding to personal identity thought experiments), and in relation to the attempts to minimize distorting influences on those reactions (recall the possible subjective influences on responses to thought experiments regarding personal identity and ethics).

²⁸³ Hare [1989], 44.

The examination of responses to thought experiments is related to grounding. From the thought experiment's set-up together with the responses it engenders, we arrive at the result of the thought experiment. In order to be more likely to obtain useful results, we need to ground the concepts and propositions of the thought experiment's set-up in the way described in Chapter 3, and ground our responses in ways such as those described in this chapter.

7. Eliciting Grounded Responses

On Unger's account, thought experiments can fail in one of two ways: first, unconscious psychological tendencies can bias responses to thought experiments. Second, thought experiments can be badly constructed. Unger's two problems are related in that badly constructed thought experiments are more likely to elicit ungrounded responses to thought experiments through interference from counter-productive psychological factors.

In this chapter, I consider how we can construct thought experiments in such a way that they are more likely to be grounded and more likely to avoid distorted responses such as those described in the previous chapter. Specifically, I give two suggestions for thought experiment construction. One concerns sufficient detail and the other the comparative method.

7.1 Stories and Details

Barbara Massey writes that successful thought experiments are well-told stories which are cogent because they have been fleshed out in compelling detail. If the thought experiment is too schematic, virtually any state of affairs can be envisioned. T. Horowitz and G. Massey summarize Barbara Massey's view as follows:

It is precisely the fine-grained articulation of an envisioned scenario, the provision of relevant and abundant detail, that has the power to rationally convince people of the possibility of some conjured-up state of affairs. A good thought experiment is just a good detailed story.²⁸⁴

I do not hold that a good thought experiment is just a good detailed story, but I do think that relevant detail is necessary to a good thought experiment. Sorensen claims that thought experiments use devices just as laboratory experiments do. A thought experimenter must be familiar with the uses of these devices and the difficulties they

²⁸⁴ Horowitz and Massey [1991], 9.

might incur. Some of the devices Sorensen lists are, "freak accidents (to eliminate characters), evil twins (to explain away damning evidence) and dream sequences (to undo previous narration)." ²⁸⁵ In this chapter, I examine devices such as these in the hope that understanding them can assist in their use in grounding thought experiments. In the course of this examination, we see that detail can be misleading or helpful. We need to be able to judge when it is serving which function.

Like Massey, Unger states that thought experiments must not be sparse and incomplete. Experiments that lack rich detail leave respondents with no firm basis from which to respond, or encourage irrelevant guesswork. ²⁸⁶ Rich detail that is relevant is helpful, for through considering the details one can ascertain more clearly whether or not the thought experiment employs assumptions that are warranted given its theoretical context.

While lack of detail may encourage irrelevant guesswork, rich detail that is irrelevant may do the same. In *Elbow Room* ²⁸⁷, Dennett examines the role that thought experiments have played in debating the problem of free will. On his account, thought experiments, which he calls "intuition pumps", can change the problem under examination. The result is that the thought experimenter is encouraged to consider the problem of free will by considering another problem, an imaginary one that is loosely related. Instead of arguing about free will, the thought experimenter has merely changed the subject. He writes:

One of my themes will be that the "classic", "traditional" free will problem of philosophy is far more an artifact of traditional methods and preoccupations of philosophers than has been recognized. ²⁸⁸

²⁸⁵ Sorensen [1992a], 243.

²⁸⁶ Unger [1990], 9.

²⁸⁷ Dennett [1984].

²⁸⁸ Dennett [1984], 6.

The traditional methods include thought experiments, or intuition pumps, which introduce problems which capitalize on fears of such things as being imprisoned, being controlled by another, or being simple automata.²⁸⁹ The thought experimenter is led to react to the fears of what is described in the thought experiment rather than to aspects of the problem of free will. For example, Dennett notes that while many hold that to have free will means that one "could have done otherwise," and many have tried to tease out what this could mean,

surprisingly little attention has been given to the question of why anyone should care about this metaphysical might-be — aside from reminding the reader that if it weren't true, why, that would be like being in prison, being paralyzed, hypnotized, a wasp, a puppet, a plaything. The allusions to the awful alternative are sometimes so swiftly traversed that quite obvious incoherence is overlooked — incoherence that would never survive the careful attention philosophers devote to their theorizing proper.²⁹⁰

Dennett does not think that intuition pumps are bad things, just that they should be used with care. He holds, as I do, that an examination of the thought experiment as an argument is a good way to check its strength.²⁹¹

So, although detail is good for clearly articulating the thought experiment and ascertaining its relationship to its theoretical context, the detail can be misleading if it leads us to consider a scenario quite distinct from that characterizing the subject under examination. In this light, consider Ayer's thought experiment:

... let us suppose that the theory of conditioning were developed to a point where it becomes possible to implant desires and beliefs and traits of character in human beings, to an extent that it could be deduced, at least in fairly general terms, how any person who had been treated in this way would most probably behave in a given situation, and that we lived under a regime in which these powers were exercised upon us, let us say from early childhood.²⁹²

²⁸⁹ Recall the sphex thought experiment in §2.5.2.

²⁹⁰ Dennett [1984], 16-17.

²⁹¹ Dennett [1984], 18.

²⁹² Ayer [1980], 9.

Ayer comments on his thought experiment in this way: "It would be as if we were spectators of a play in which we also participated, with no other option than to enact the roles allotted to us."²⁹³ Dennett responds:

But wait just a minute; suppose we began to giggle and whisper, and completely disrupted the play by pointing out to each other how pointless the plot was. Unfair to Ayer's example! We are not supposed to imagine ourselves being able to do that. But why not? Did Ayer stipulate that in this imaginary world people can no longer notice things, and react according to them? The evil scientists who run this world can apparently peer through the armored glass at us without worrying that we will unexpectedly shake our fists at them and revolt. It must be because, without saying so, Ayer has turned us into much simpler creatures.²⁹⁴

Dennett's criticism is that this thought experiment oversimplifies a situation by minimizing the great differences between us and these programmed automata. It lures us into considering that we might indeed be in this type of a situation, and so we think that the reactions we have are pertinent to the free will problem. In this case, both the lack of detail and the irrelevant details have led us to consider that an oversimplified situation is analogous and pertinent to the one we are considering.

In an exchange between Shoemaker and Unger, two story-like thought experiments are posed about an imaginary place called Brainland. By slightly changing the details of Shoemaker's thought experiment, Unger, who is a physicalist, creates a thought experiment that seems to yield very different responses from those of Shoemaker. Unger states his physicalist position in the following way:

For you to exist at some particular future time, there must be the sufficiently continuous physical realization of a core psychology between the physical realizer of your core now (your brain) and the physical realizer (whether brain or not) of someone's core psychology at that future time.²⁹⁵

²⁹³ Ayer [1980], 10.

²⁹⁴ Dennett [1984], 34.

²⁹⁵ Unger [1992a], 134.

For Unger, core psychology consists of things like the capacity for conscious experience and the capacity for very simple reasoning.²⁹⁶ Distinctive psychology seems to include all other aspects of one's dispositional psychology. Some of these capacities are shared with others (for example, one's "(ostensible) memory of having tasted butter pecan ice cream"²⁹⁷) and some are not (for example, one's memory of uniquely individual experiences²⁹⁸).

In critiquing Unger's view that the continuity of core psychology, rather than distinctive psychology, is necessary for survival, Shoemaker presents the following thought experiment about Brainland: In Brainland, brains are anchored, one per square foot, to life support systems. The brains communicate electronically with other brains. Brains are sometimes dormant and sometimes active. When a brain wishes to move, it locates a dormant brain beside it and moves its distinctive psychology to that brain. Shoemaker claims that persons go with their distinctive psychology, "strolling" from brain to brain, rather than staying with the core psychology in a particular brain.²⁹⁹

Unger claims that we are used to suspending disbelief in science-fiction examples for the sake of the story. He charges that the story-like quality of Brainland — rather than liberating us from notions that may be incorrect — predisposes us to view the thought experiment uncritically and so to arrive at the wrong response. Contrary to Shoemaker, Unger holds that persons in Brainland stay with their core psychologies. Unger defends this response with his own story of Lag Brainland where the distinctive psychology is shared for a few seconds by the original brain and the new brain. Then,

²⁹⁶ Unger [1992a], 134.

²⁹⁷ Unger [1990], 68.

²⁹⁸ Unger [1990], 68.

²⁹⁹ Shoemaker [1992], 141.

when the original brain loses the distinctive psychology, Unger claims we hold that this resulting unfortunate moron is the surviving person.³⁰⁰

Thought experiments such as Brainland and Lag Brainland are misleading, not necessarily because they predispose us to go along with all sorts of impossibilities for the sake of the story, such as assuming character continuity (perhaps none of the brains carry the identity), but because the choice of the details that are to be included or left out of the thought experiment can inappropriately influence one's responses. We see this in the different responses given to Shoemaker's and Unger's thought experiments.

Presentation matters. This is so, even in physical experiments. In the following passage, Gooding draws our attention to the mediating role of representations. He writes:

I have argued that what scientists do when "intervening" shapes how they represent the outcomes of their interventions. Successful representation in turn influences how they theorize those outcomes. Experimental practices are sometimes essential to the construction of ways of experiencing and theorizing new phenomenal domains, although they are not the only resource for such constructs.³⁰¹

In scientific experiments, representations are chosen to present the natural phenomena under study. They aid understanding and disseminating of information. In thought experiments the role of representations is also very important.

I have noted that thought experiments contain tacit *ceteris paribus* clauses.³⁰² It is always possible to add more details to thought experiments, but the clause is to be understood as meaning that other possible details do not affect the variables in question.³⁰³ It is important to examine the variables and try to ascertain which details are

³⁰⁰ Unger [1992b], 162.

³⁰¹ Gooding, Pinch and Schaffer [1989], 215.

³⁰² See §3.2.

³⁰³ Sorensen [1992a], 247.

relevant to them and which are not, in order to use the *ceteris paribus* assumption successfully.

When a thought experiment is set out in sufficient detail it is easier to ascertain whether the question it poses is well-formed, whether it uses concepts that are defined normally, whether it rests on unwarranted assumptions, and whether it is inconsistent with the theoretical context. However, care must be taken. Details can help one examine and clarify assumptions and their relation to the theoretical context. At the same time we have seen in the Brainland thought experiments that careful choosing of detail can unduly predispose responses favourable to certain views. The comparative method, which I outline in §7.3, helps to mitigate against the swaying of intuition through carefully selected detail.

7.2 Contextualism, Literature and Detail

The contextualist approach to ethics and Nussbaum's insight into Jamesian literature can help with our understanding of thought experiments, for both approaches focus on the merits of considering the specific detail of a particular situation in the making of ethical decisions. Here, I examine the manner in which my account of grounding relates to Nussbaum's interpretation of James' *The Golden Bowl*.

Literature is more overtly descriptive than thought experiments, and thought experiments are more overtly purposeful than literature. While it is often the purpose of literature to spark an emotional response, to be expansive, concrete and rich, it is often the goal of thought experiments to encourage an impartial view that is reductionistic, general and abstract.

Nussbaum stresses that it is through detail that James grounds his story in the objective, human social experience which Nussbaum claims is both an interpretation and a measure of existing value. For Nussbaum, existing value is measured by an internal human objectivity, which is not an objectivity with respect to the external world, but

rather with respect to the history of human social experience. This history both interprets and measures norms of rightness. Moral perception is not a perceiving of the "world apart from human beings and human conceptual schemes, but a world already interpreted and humanized by our faculties and concepts."³⁰⁴ Just as some people have better eyesight and hearing, so some people have better moral perception.

Characteristics that Nussbaum claims make *The Golden Bowl* a work of moral philosophy are some of the same characteristics that contribute to a successful thought experiment. In particular, the attention to the perception of finely tuned detail and nuance creates literature and thought experiments that exhibit the characteristics that allow us to decide if the literature or thought experiment is adequately related to its theoretical context.

Nussbaum claims that Jamesian literature is both a moral act and a work of moral philosophy. It is difficult to establish that any literature is a moral act or a work of moral philosophy. At any rate, this claim would depend on how we define the moral enterprise. However, it does seem safe to claim that some literature, like some thought experiments, leads us to new conclusions about the questions it explores in the moral domain.

Perhaps it is in the similarities between Nussbaum's conception of Jamesian literature and my conception of thought experiments that the most insight can be gained. One possible insight is similar to the one that the contextualist brings to ethics, namely, every situation, due to its particular detail, suggests new and unpredictable resolutions to the question being studied. Through attention to context, successful thought experiments can be a methodological link between the overly reductionistic models of theoretical ethics and the detail of real life.

³⁰⁴ Nussbaum [1989], 131.

The experiment part of a thought experiment transforms it from being solely a case of hypothetical reasoning to something more useful and grounded in the observable world. The finely realized detail of some literature, that is responsive to moral situations in their particular contexts, may make that literature useful as moral philosophy.

The fact that we can see that there is a connection between a controlled experiment and the detail and sensitivity of art is interesting. The connection is found through the faculties of imagination and perception which discover the surprising, complex and finely-tuned details of each particular situation.

Nussbaum's work suggests a refinement to the answer given earlier to the concern about whether thought experiments can tell us anything new about the world. I stated that thought experiments expose implicit beliefs and so reveal false or inconsistent background theories. Nussbaum's work points to how these implicit beliefs surface. She would perhaps claim that it is through specifying the detail of the context of the unique situation that thought experiments in ethics expose something new about the experienced human social world.

7.3 Comparative Method

I call the examination of sets of related thought experiments "the comparative method". Consider Thomson's use of the comparative method. We have seen that in Thomson's ill-violinist thought experiment she is testing the question "Is abortion permissible?" by posing the analogous question "Is disconnecting permissible?" Here, imagination is used in two ways. First, it is used to create an analogous situation that is less burdened by preconceived notions and prejudices than the main situation under examination. Second, it is used to vary the conditions under which a specific question is tested. Remember, Thomson has stated that you have been kidnapped and plugged into an ill violinist. She then asks you to imagine a variety of cases with the following features:

- a. You are plugged in for 9 years.³⁰⁵
- b. You'll only live for a month, because the strain of the violinist using your kidneys will kill you.³⁰⁶
- c. The violinist needs only one hour of your life and letting him use your kidneys for that one hour would not affect your health.³⁰⁷

Thomson uses these variations of her original thought experiment in order to make a distinction between "right" and "ought". Rights, Thomson remarks, cannot be altered in the event that fulfilling them is either easier or harder. It does not make sense to say that one has a right to something when it is easy to fulfill, but not when it's difficult.

She concludes:

So my own view is that even though you ought to let the violinist use your kidneys for the one hour he needs, we should not conclude that he has a right to do so — we should say that if you refuse, you are, like the boy who owns all the chocolates and will give none away, self-centred and callous, indecent in fact, but not unjust. And similarly, that even supposing a case in which a woman pregnant due to rape ought to allow the unborn person to use her body for the hour he needs, we should not conclude that he has a right to do so; we should conclude that she is self-centred, callous, indecent, but not unjust, if she refuses ... If anyone does wish to deduce "he has a right" from "you ought", then all the same he must surely grant that there are cases in which it is not morally required of you that you allow the violinist to use your kidneys, and in which he does not have a right to use them, and so also for mother and unborn child.³⁰⁸

A particular thought experiment may be considered in a variety of contexts and with a variety of perspectives and details. In comparing and checking for inconsistencies amongst the resulting responses, the responses may be considered in relation to the

³⁰⁵ Paraphrased from Thomson [1971], 46.

³⁰⁶ Paraphrased from Thomson [1971], 42.

³⁰⁷ Paraphrased from Thomson [1971], 49.

³⁰⁸ Thomson [1971], 51.

question of the thought experiment in order to determine which present philosophically interesting challenges and why.

Williams gives an account of thought experiments that also suggests facilitating the grounding of responses by comparing various related thought experiments. Williams presents the following series of cases. We are to imagine that we are the person referred to as "A" and that after one of the procedures listed below, A is to be tortured:

- (i) A is subjected to an operation which produces total amnesia;
- (ii) amnesia is produced in A, and other interference leads to certain changes in his character;
- (iii) changes in his character are produced, and at the same time certain illusory "memory" beliefs are induced in him; these are of a quite fictitious kind and do not fit the life of any actual person;
- (iv) the same as (iii), except that both the character traits and the "memory" impressions are designed to be appropriate to another actual person, B;
- (v) the same as (iv), except that the result is produced by putting the information into A from the brain of B, by a method which leaves B the same as he was before;
- (vi) the same happens to A as in (v), but B is not left the same, since a similar operation is conducted in the reverse direction.³⁰⁹

I introduce this series of cases to show how detail added or left out of a thought experiment makes substantial differences to our responses. If Williams had only supplied the thought experiment outlined in (vi), one might be inclined to go with the idea that one's identity would switch bodies. One might then feel terrible that B would be tortured, but relieved to think that one would be safe in B's body. This conclusion is consistent with a psychological/distinctive psychology view of personal identity. However, on considering cases (i) through (v), one would be less sure of the conclusion about (vi). After considering all the cases, it seems that one would worry about the torture regardless of the content of one's memories. The new position one might be inclined to accept is a physicalist/core psychology view. In Williams' series of cases, we see that the cases considered together elicit a more refined responses than any one of them alone might.

³⁰⁹ Williams [1970], 190.

I suggest that Williams' strategy could be used as a test for sufficient detail in a thought experiment. In constructing a thought experiment, it would be helpful to see if there are a number of ways to fill in its details, and to examine whether the details change the responses to the thought experiment. If they do, then it would be necessary to examine why, and to examine the tension between the various cases and the responses. It may be that, as in Williams' case, the tension is not easy to resolve.

Williams' set of strategies suggest ways to minimize psychological distortions. In these cases, Williams adds the fact that A is to be tortured after the procedure described in order to focus the response. The response might be more casual without the consideration of oneself as the recipient of the torture. This is a similar strategy to that of Unger in his AFGP tests. Of this, Williams writes:

If I expect that a certain situation, S, will come about in the future, there is of course a wide range of emotions and concerns, directed on S, which I may experience now in relation to my expectation. Unless I am exceptionally egoistic, it is not a condition on my being concerned in relation to this expectation, that I myself will be involved in S — where my being "involved" in S means that I figure in S as someone doing something at that time or having something done to me, or, again, that S will have consequences affecting me at that or some other subsequent time. There are some emotions, however, which I will feel only if I will be involved in S, and fear is an obvious example.³¹⁰

Dennett also recommends varying thought experiments and tracking the changes in responses as a way to assess thought experiments. His examples are designed to show how changing the details of a thought experiment can radically alter one's response. Understanding the different responses is helpful in clarifying the problem posed by the thought experiment. Here are the three examples Dennett offers:

Consider the frequently discussed case of the demonic neurologist who directly manipulates a person's brain to induce all his desires, beliefs, and decisions.³¹¹

³¹⁰ Williams [1970], 193.

³¹¹ Fischer [1982], 37. Given in Dennett [1984], 65.

Consider the infrequently discussed case of the eloquent philosopher who indirectly manipulates a person's brain by bombarding his ears with words of ravishing clarity and a host of persuasively presented reasons, thereby inducing all his desires, beliefs, and decisions.

Consider the delightful case of the well-informed, truthful oracle who indirectly manipulates a person's brain by bombarding his ears with lucid and accurate warnings, made all the more irresistible by the citation of all the evidence in their favor and a frank account of the entire evidence-gathering operation.³¹²

First, note that these examples are not thought experiments, but rather invitations to examine a certain state of affairs. However, they make the point that differences in presentation often do result in differences in responses. Dennett uses the different responses to explore what it means to be influenced in the various cases and what repercussion that has for questions of free will. Here, however, it is enough to note that the various presentations do seem to influence our responses and that Dennett's strategy of running a series is a good one. Through examining the series we can examine which features of the presentations make the responses vary and why.

Thought experiments often take on the form of conversations in which each new experimenter takes the thought experiment and modifies it. This is a kind of comparative method. Hacking claims that thought experiments do not have a life of their own, in that they are "rather fixed, largely immutable."³¹³ Of physical experiments he writes, "I think of experiments as having a life: maturing, evolving, adapting, being not only recycled but also, quite literally being retooled."³¹⁴ However, thought experiments do evolve, just as physical experiments do. In the course of being discussed they are varied and refined. Consider Unger and Shoemaker's development of the Brainland thought experiment given above, or consider the Hobbes' addition to Plutarch's Ship of Theseus thought experiment given in §5.2.

³¹² The last two versions are from Dennett [1984], 64-65.

³¹³ Hacking [1993], 303.

³¹⁴ Hacking [1993], 303.

Some have argued that the fact that our responses to thought experiments change in relation to presentation makes the thought experiment methodology suspect. Unger notes that our responses to thought experiments are affected by their proximity to other thought experiments and by the social context in which they arise.³¹⁵ The main point of this chapter is that this is not entirely unhealthy. We can use the comparative method to try to elicit contradictory responses in order to determine what it is about the method of presentation that affects the responses.

Even so, a problem with the comparative method is raised by Horowitz. In examining the following pair of cases, presented by Quinn, she argues that it is differences in the reasoning process, rather than moral differences related to the killing and letting die distinction that influence people's responses. Here are the thought experiments:

Rescue Dilemma 1: We can either save five people in danger of drowning in one place or a single person in danger of drowning somewhere else. We cannot save all six.

Rescue Dilemma 2: We can save the five only by driving over and thereby killing someone who (for an unspecified reason) is trapped on the road. If we do not undertake the rescue the trapped person can later be freed.³¹⁶

This thought experiment series is designed to elicit responses that give insight into the difference between doing and allowing harm to another. Horowitz presents the view that this thought experiment series leads one to choose a different neutral outcome for each scenario and so the response is not about the moral differences between the actions presented. She claims that the way people weigh outcomes is more closely described by Kahneman and Tversky's prospect theory than by classical expected-utility theory.³¹⁷

³¹⁵ Unger [1982], 120.

³¹⁶ Quinn [1993], 149. Given in Horowitz [1998], 368.

³¹⁷ Horowitz [1998], 369.

Furthermore, since this reasoning process is unconscious, the experimenter has no way of knowing that their response is due to reasoning processes and does not reflect moral preferences. Horowitz concludes that

If this is right, then one's intuition that there is this difference in the force of reasons would not be explained in terms of a perceived difference between action and inaction but rather in terms of differing responses to gains and losses.³¹⁸

There are two responses to Horowitz's account. First, comparable cases need to be comparable in the sense that they isolate the same variable. Her case rests on the fact that Quinn's two thought experiments are not variations of the *same* experiment. As she observes, each of the two experiments sets reasoning in process that is relative to a different neutral outcome. However, her account does not affect my claim that it is a good thing to compare versions of a thought experiment which are the same in all respects, save for one carefully controlled variation, in the manner of Williams and Thomson. Secondly, Horowitz's account points to the necessity of understanding our reasoning processes as well as possible and trying to make them conscious so that our reactions to thought experiments are not affected by the manner of reasoning employed, but rather by the variable we wish to track.

In this section, we have seen how Unger, Thomson and Williams employ sets of thought experiments to test questions across a variety of situations. The results are compared and related to the variations in the thought experiments. This method results in more clearly defined questions, more explicit theoretical contexts and more grounded responses. Like refinements of practical inventions that lead to new developments in theory, refinements of thought experiments lead to clearer and more developed theories. For example, in the Thomson experiments we see how the original thought experiment explored the balance of the right to life of a foetus with the right of the mother to

³¹⁸ Horowitz [1998], 378.

determine what happens in and to her body. The subsequent variations of the thought experiment develop the difference between unjust and indecent behaviour. In the Williams thought experiments we see how the detail supplied led us to examine the relationship between physical and psychological continuity and then to the refinement of the idea of core and distinctive psychological continuity and its supervenience on the physical.

Sorensen writes that biases to thought experimentation need to be considered, but do not present an insurmountable obstacle. We are both susceptible to bias and adept at correcting bias.³¹⁹ For guarding against bias in thought experiments he suggests five strategies.³²⁰ First, he suggests that others be asked for their reactions to thought experiments, particularly those who hold adverse theories. Second, he suggests using variations of the thought experiments. Third, he suggests using old thought experiments that were discussed prior to one's theory, and comparing old intuitions to current ones. Fourth, he suggest introducing a bias against the hypothesis that cancels out one's bias. Finally, he suggests gathering

wild specimens of intuitions. For example, a thought experiment mounted in support of the possibility of unfelt pain can be corroborated with medical texts describing operations that "block pain" and a line from Captain Horatio Hornblower, in which C.S. Forester writes, "Hornblower found the keen wind so delicious that he was unconscious of the pain the hailstones causes him."³²¹

This chapter was written with the same purpose as Sorensen's suggestions. We rely on our responses to thought experiments, but we can construct our thought experiments to purposefully mitigate against experimenter biases affecting those responses.

³¹⁹ Sorensen [1992a], 262.

³²⁰ Sorensen [1992a], 263.

³²¹ Sorensen [1992a], 263.

I have been concerned, in this chapter, with the tests and methods we can employ in order to minimize distorted, ungrounded responses to thought experiments. My approach has similarities to the way that one checks for accuracy in any scientific process. Consider microscopy. Here, we want to be able to say that the image is a true one. We do this by double-checking and seeing past the distortions. Here is an example of double-checking an image seen through a microscope:

The conviction that a particular part of a cell is there as imaged is, to say the least, reinforced when, using straightforward physical means, you microinject a fluid into that part of the cell. We see the tiny glass needle — a tool that we have ourselves hand crafted under a microscope — jerk through the cell wall. We see the lipid oozing out the end of the needle as we gently turn the micrometer screw on a large, thoroughly macroscopic, plunger.³²²

If we use two different techniques to observe the same body we can say with some confidence that the body exists. Hacking writes that we are convinced about the structures we see,

not by a high powered deductive theory about the cell — there is none — but because of a large number of interlocking low level generalizations that enable us to control and create phenomena in the microscope.³²³

How do we do these double-checks in thought experiments? In ethics we might use a thought experiment to come to a certain conclusion about a particular behaviour in a particular situation. Then, we need to see if it holds in a variety of situations. When we use the comparative method we feel more confident that we have a grounded response when a variety of thought experiments elicit the same response. When the compared thought experiments do not, we are led to ask why the response is different. This is Unger's strategy in the shallow pond and the envelope thought experiment given in §2.5.2. Moreover, the comparative method is useful in refining the results of thought experiments. For example, through the comparative method, we may see the limits of

³²² Hacking [1983], 190.

³²³ Hacking [1983], 209.

applicability of a particular concept or idea. For example, in considering the two presentations of Brainland we see the limits of applying the criterion of psychological continuity to personal identity.

Microscopists also have to learn to see past distortions. For example, with an interference-contrast microscope perceived objects have asymmetrical fringes around them that the microscopist learns to disregard.³²⁴ In thought experiments too, we need to recognize and account for distortions to our responses. In the previous chapter, I outlined possible distortions to responses from lack of individual understanding of the theoretical context, lack of development of the theoretical context, and psychological and subjective tendencies. In this chapter, we see that adding details can lead us to entertain certain responses over others, and that the best strategy for grounding our responses is to compare various relevantly detailed thought experiments. Through the details, we can both ascertain whether the thought experiment's set-up is grounded or not, by understanding more clearly its relationship to its theoretical context, and also form more grounded responses to the thought experiment. This gives us a higher probability of interesting results.

³²⁴ Hacking [1983], 191.

8. Contextualism and Thought Experiments

I began this dissertation with the questions: "How can an experiment which occurs in thought lead to new and accurate conclusions about the world outside thought?" and, "What makes thought experiments relevant to the domain they are designed to explore?" The first step in answering these questions was given, in Chapter 2, by an analysis of the features that thought experiments share with physical experiments. A thought experiment, like a physical experiment, isolates and varies variables in order to answer a question within a given theoretical context. The result of the experiment has repercussions for that theoretical context.

I distinguished between the context of discovery and the context of justification of a thought experiment. In the context of discovery, the experiment is created; this process may be as diverse as any creative process. In the context of justification, the reasoning employed in the thought experiment and the relationship of the thought experiment to its theoretical context is assessed; this process has discernible steps, which I outline in this dissertation.

Before the grounding of the thought experiment itself can be considered the reasoning employed in the thought experiment needs to be analyzed. I claim that thought experiments must be amenable to reconstruction as arguments. These arguments can be examined for deductive validity or inductive strength. Many thought experiments employ reasoning by analogy. Because assessing analogies is a difficult enterprise, I analyze possible difficulties with analogies and ways to anticipate or discover them.

Some thought experiments test the explanatory potential of a theory. For example, thought experiments such as Darwin's in *On the Origin of Species* do just this. The explanation given by this type of thought experiment needs to be assessed. It must be consistent with the relevant background theories and plausible with respect to the

evidence. Then, the explanation needs to be ranked in relation to other possible explanations.

Some types of computer simulations are like thought experiments. I call these simulation experiments. Simulation experiments either test the explanatory potential of a theory or test causal links between one state of affairs and another. The flocking simulation tests the explanatory potential of a theory; Danielson's PD simulations test the links between one state of affairs and another. Just as some thought scenarios, such as Maxwell's Demon,³²⁵ are not thought experiments, so, some simulations are not simulation experiments. I call these "simulation demonstrations". Flight simulators are used as tools to understand and interact with modeled states of affairs that are constructed from well-established theories. They do not test the explanatory potential of a theory, or test links between one state of affairs and another.

I have given an account of grounding for both the experimental set-up of the thought experiment, and for responses to the thought experiment. With respect to the former, the concepts used must be defined normally, dependent and independent variables must be isolated and relevantly related, and the propositions of the thought experiment (excepting those describing extraneous particulars) must be relevantly related to the given theoretical context and the question under examination. With respect to the latter, responses to the thought experiment must arise from a good knowledge of the theoretical context, from a clearly defined theoretical context and from an awareness of the influence of possible subjective distortions.

Throughout the dissertation, I have presented examples of thought experiments from various fields such as ethics, science, metaphysics and philosophy of language in order to illustrate my account.

³²⁵ See §2.6.

In Chapter 4, I turned to the examination of thought experiments concerning ethics. Contextualism supports my account of the role and efficacy of thought experiments in ethics. The contextualist approach also suggests a framework from which to assess the grounding of these thought experiments. Consider Thomson's views of the importance of considering situational aspects of ethical problems. She writes,

But the fact is that there is no end to the range of possible situations in which human beings may find themselves, or of the range of beliefs, intentions, and motives they may act on in those situations. If we need no special equipment to observe the data, we do all the same need to remind ourselves from time to time of the wide variety of possible human action, and of the fact that a wide array of considerations bear on all but the simplest of moral problems.

This comes home to us particularly vividly when we turn to the literature of law. Case books are like anthologies of short stories, each of which ends in a moral problem. Defendant Smith was driving his car down Main Street when a child ran out from between two parked cars, and to avoid him, Smith swerved. Unfortunately, Smith had not been watching quite as carefully as he should have been, so he swerved to the right instead of to the left — and drove smack into Plaintiff Jones. Should Defendant Smith be placed under the duty of paying Plaintiff Jones' medical bill? If so, why? If not, why not? The judge's answer is constrained in ways in which the moral theorist's is not; but the moral theorist can learn, not merely from the story itself, that is, not merely from the possibilities of human action which the story reminds us of, but from the judge's decision and his or her argument for it.

Many moral theorists came to the literature of law by way of what seems to me the most interesting work ever written on its topic, namely *Causation in the Law*, by H.L.A. Hart and A.M. Honore.³²⁶ No moral theorist, however, ingenious, could have invented the stories which Hart and Honore told us about what actual people actually did to each other. Contact with law has been immensely enriching to moral theory in recent years.³²⁷

Like the examples from the literature of law, thought experiments give us details. They aid us in envisioning how our moral theories and frameworks of values might be acted out, or relate to what is acted out. Through thought experiments we can see how beliefs about our values work in relation to particular situations. The term casuistry, which is used to describe the practice of case-by-case reasoning, has been used in the modern

³²⁶ Hart and Honore [1959].

³²⁷ Thomson [1986], 256-257.

period mainly in the pejorative sense, since those who employ this type of reasoning are seen as abandoning principles for case-specific considerations. However, the contextualist process uses cases in a more developed way. Principles are not abandoned, but considered along with a variety of other concerns. The principles-only approach to ethical problems is just as impoverished as the situations-only approach. In the wide reflective equilibrium process constraints in operation on a particular situation are weighed. The factors considered in the wide reflective equilibrium process — possible judgements and outcomes, beliefs arising from the framework of values, moral principles, background theories and situational factors — are just those which we need to consider in judging whether a thought experiment is grounded. These factors are the main features of primary and background parts of the theoretical context to which the thought experiment must be in a grounded relation.

In Chapter 7, I suggest ways to construct thought experiments that have a higher likelihood of being grounded. My strategy is similar to that outlined in the following passage:

Perceptual psychologists like to characterize vision as a "bag of tricks." We manage to make sense of what we see by employing a large group of overlapping rules of thumb. Each rule has its weaknesses which can be isolated in a contrived setting. But when operating collectively, as they normally do, the rules are reliable because there are few situations in which they go wrong simultaneously and in the same direction. The lesson here is that we must avoid the fallacy of composition: from the unreliability of the parts, we cannot infer the unreliability of the whole.³²⁸

My "bag of tricks" includes creating a thought experiment that has a good experimental structure (isolates the variables and uses simplifying assumptions properly), employs strong reasoning (has good analogies, valid deductive or strong inductive reasoning, or suggests a good explanation), is adequately related to its theoretical context (is grounded in that context) and employs considered responses. The comparative

³²⁸ Sorenson [1992b], 36.

method I suggest in Chapter 7 facilitates the constructing of grounded thought experiments, and the assessing of the grounding of thought experiments. It also facilitates incurring grounded responses to thought experiments. It does this by supplying the relevant detail through which the relationship of the thought experiment to its theoretical context can be adequately examined.

Finally, note that my approach is reliabilist. I am seeking a way to construct thought experiments that are reliably useful to us and relevant to our beliefs about values. Consider the following passage by Hacking:

There are surely innumerable entities and processes that humans will never know about. Perhaps there are many that in principle we can never know about. Reality is bigger than us. The best kinds of evidence for the reality of a postulated or inferred entity is that we can begin to measure it or otherwise understand its causal powers. The best evidence, in turn, that we have this kind of understanding is that we can set out, from scratch, to build machines that will work fairly reliably, taking advantage of this or that causal nexus. Hence, engineering, not theorizing, is the best proof of scientific realism about entities. My attack on scientific realism is analogous to Marx's onslaught on the idealism of his day. Both say that the point is not to understand the world but to change it.³²⁹

Hacking's instrumental realism about science is similar to mine about ethics. Reliable moral approaches are those that adequately aid in fulfilling the design criteria relating to the problem at hand. The approach can be justified because it works reliably — it strikes an acceptable, and so useful, balance between the multiple constraints acting on each moral situation.

³²⁹ Hacking [1983], 274-275.

Bibliography

- Arbib, Michael A. [1977] *Computers and the Cybernetic Society*, second edition, Orlando: Academic Press, 1984.
- Aristotle, *Nichomachean Ethics*, David Ross (trans.), Oxford: Oxford University Press, 1991.
- Axelrod, Robert [1984] *The Evolution of Cooperation*, New York: Basic Books.
- Ayer, A.J. [1980] "Free-will and Rationality", in Zak van Straaten (ed.), *Philosophical Subjects: Essays Presented to P.F. Strawson*, Oxford: Clarendon Press, 1-13.
- Bailey, James [1992] "First We Reshape Our Computers, Then Our Computers Reshape Us: The Broader Intellectual Impact of Parallelism", *Daedalus* 121, 67-86. Cited from the abridged and reprinted paper "New Computer, New Thoughts", in Kathryn Schellenberg (ed.), *Computers in Society*, fifth edition, Guilford: Dushkin Publishing, 1994, 232-235.
- Baier, Annette [1987] "Hume, the Women's Moral Theorist?", in Eva Feder Kittay and Diane T. Meyers (eds) *Women and Moral Theory*, Totowa, NJ: Rowman and Littlefield, 37-55.
- Binmore, Ken [1994] *Game Theory and the Social Contract*, Volume I: *Playing Fair*, Cambridge: MIT Press.
- Bohr, N. [1949] "Discussion with Einstein on Epistemological Problems in Atomic Physics". Cited from P.A. Schlipp (ed.), *Albert Einstein Philosopher-Scientist*, La Salle: Open Court, 1959, 199-241.
- Boniolo, Giovanni [1997] "On a Unified Theory of Models and Thought Experiments in Natural Sciences", *International Studies in the Philosophy of Science* 11, 121-142.
- Brown, J. [1993a] "Author's Response", *Metascience: New Series* 3, 38-40.
- Brown, J. [1993b] "Why Empiricism Won't Work", in David Hull, Micky Forbes and Kathleen Okruhlik (eds), *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*, Vol. 2, East Lansing, MI: Philosophy of Science Association, 271-279.
- Brown, J. [1991a] *The Laboratory of the Mind*, London: Routledge.
- Brown, J. [1991b] "Thought Experiments: A Platonic Account", in Tamara Horowitz and Gerald J. Massey (eds), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 119-128.
- Bunzl, Martin [1995] "Bunzl on Sorensen's *Thought Experiments*", *Informal Logic* 17, 388-393.
- Burkholder, Leslie [1998] "Killing, Letting Die, and Rachels's Thought Experiment", unpublished.

- Cole, David [1984] "Thought and Thought Experiments", *Philosophical Studies* 45, 431-444.
- Coombs, Jerrold R. and Earl R. Winkler (eds) [1993] *Applied Ethics: A Reader*, Cambridge: Blackwell.
- D'Agostina, Fred [1997] "Original Position", *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/orginal-position/>, first posted: February 27, 1996. Cited from last modification: July 28th, 1997.
- Dancy, Jonathan [1985] "The Role of Imaginary Cases in Ethics", *Pacific Philosophical Quarterly* 66, 141-153.
- Daniels, N. [1979] "Wide Reflective Equilibrium and Theory Acceptance in Ethics", *Journal of Philosophy* 76, 256-82.
- Danielson, Peter [1995a] "Evolving Artificial Moralities: Genetic Strategies, Spontaneous Order, and Moral Catastrophe", in A. Albert (ed.), *Chaos and Society*, Amsterdam: IOS Press, 293-308.
- Danielson, Peter [1995b] "Prisoner's Dilemma Popularized: Game Theory and Ethical Progress", *Dialogue* 34, 295-304.
- Danielson, Peter [1992] *Artificial Morality*, London: Routledge.
- Darwin, Charles [1859] *On the Origin of Species*. Cited from Darwin, *On the Origin of Species: A Facsimile of the First Edition*, New York: Atheneum, 1967.
- Dennett, Daniel C. [1987] *The Intentional Stance*, Cambridge: MIT Press.
- Dennett, Daniel C. [1984] *Elbow Room*, Cambridge: MIT Press.
- Doran, Jim and Nigel Gomila (eds) [1994] *Simulation Societies: The Computer Simulation of Social Phenomena*, London: UCL Press.
- Duhem, Pierre [1906] *La Théorie Physique: Son Objet, Sa Structure*, second edition, Paris: Marcel Riviere & Cie, 1914. Cited from Philip P. Wiener (trans.), *The Aim and Structure of Physical Theory*, Princeton: Princeton University Press, 1954.
- Einstein, Albert [1949] "Autobiographical Notes", in P.A. Schlipp (ed.), *Albert Einstein Philosopher-Scientist*, La Salle: Open Court, 1959, 1-95.
- Euclid, *Elements*, Thomas Heath (ed.), New York: Dover, 1959.
- Feldman, Richard [1995] "Feldman on Sorensen's *Thought Experiments*", *Informal Logic* 17, 394-398.
- Fischer, J. [1982] "Responsibility and Control", *Journal of Philosophy* 79, 24-40.
- Forbes, M., David Hull and Kathleen Okruhlik (eds) [1993] *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*, Vol. 2, East Lansing, MI: Philosophy of Science Association.

- Frolich, Norman and Joe Oppenheimer [1997] "A Role for Structured Observation in Ethics", *Social Justice Research* 10, 1-21.
- Gale, Richard M. [1991] "On Some Pernicious Thought Experiments", in Tamara Horowitz and Gerald J. Massey (eds), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 297-303.
- Galileo, Galilei [1638] *Dialogues Concerning Two New Sciences*, Henry Crew and Alfonso de Salvio (trans), Evanston: Northwestern University Press, 1950.
- Gärdenfors, Peter [1988] *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, Cambridge: MIT Press.
- Gardiner, Patrick [1983] "Professor Nussbaum on *The Golden Bowl*", *New Literary History* 15, 178-84.
- Gillett, Grant [1986] "Brain Bisection and Personal Identity", *Mind* 95, 224-9.
- Gillies, Donald [1993] *Philosophy and Science in the Twentieth Century*, Oxford: Blackwell.
- Gomila, Antoni [1991] "What is a Thought Experiment?", *Metaphilosophy* 22, 84-92.
- Gomila, Nigel and Jim Doran (eds) [1994] *Simulation Societies: The Computer Simulation of Social Phenomena*, London: UCL Press.
- Gooding, D.C. [1993] "What is Experimental about Thought Experiments?" in David Hull, Micky Forbes and Kathleen Okruhlik (eds), *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*, Vol. 2, East Lansing, MI: Philosophy of Science Association, 280-290.
- Gooding, D.C. [1992] "The Procedural Turn; or, Why Do Thought Experiments Work?", in Ronald N. Giere (ed.) [1992] *Cognitive Models of Science: Minnesota Studies in the Philosophy of Science* 15, Minneapolis: University of Minnesota Press, 45-76.
- Gooding, D.C. [1990] *Experiment and the Making of Meaning*, Dordrecht: Kluwer.
- Gooding, D.C., Trevor Pinch and Simon Schaffer (eds) [1989] *The Uses of Experiment*, Cambridge: Cambridge University Press.
- Hacking, Ian [1993] "Do Thought Experiments Have a Life of Their Own? Comments on James Brown, Nancy Nersessian and David Gooding", in David Hull, Micky Forbes, and Kathleen Okruhlik (eds), *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*, Vol. 2, East Lansing, MI: Philosophy of Science Association, 302-308.
- Hacking, Ian [1983] *Representing and Intervening*, Cambridge: Cambridge University Press.
- Hare, R.M. [1989] *Essays in Ethical Theory*, Oxford: Clarendon Press.
- Hare, R.M. [1981] *Moral Thinking*, Oxford: Oxford University Press.

- Hare, R.M. [1971] *Essays on Philosophical Method*, London: Macmillan Press.
- Hare, R.M. [1963] *Freedom and Reason*, Oxford: Clarendon Press.
- Harman, Gilbert [1986] "Moral Explanations of Natural Facts — Can Moral Claims Be Tested Against Moral Reality?", *Southern Journal of Philosophy* Supp. Vol. 24, 57-68.
- Harris, John [1975] "The Survival Lottery", *Philosophy* 50, 81-7. Cited from Peter Singer (ed.), *Applied Ethics*, Oxford: Oxford University Press, 1986, 87-95.
- Hart, H.L.A., and A.M. Honore [1959] *Causation in the Law*, Oxford: Oxford University Press.
- Hertzberger, Hans [1970] "Paradoxes of Grounding in Semantics", *Journal of Philosophy* 67, 145-167.
- Hobbes, Thomas [1655] *De Corpore*. Cited from Mary Whiton Calkins (ed.), *The Metaphysical System of Hobbes in Twelve Chapters from Elements of Philosophy Concerning Body*, Chicago: Open Court, 1905.
- Hon, Giora [1989] "Towards a Typology of Experimental Errors: An Epistemological View", *Studies in History and Philosophy of Science* 20, 469-504.
- Honore, A.M., and H.L.A. Hart [1959] *Causation in the Law*, Oxford: Oxford University Press.
- Horowitz, Tamara [1998] "Philosophical Intuitions and Psychological Theory", *Ethics* 108, 367-385.
- Horowitz, Tamara and Gerald J. Massey (eds) [1991] *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield.
- Hull, David L. [1997] "That Just Don't Sound Right: A Plea for Real Examples", presented at University of British Columbia.
- Hull, David L., Micky Forbes and Kathleen Okruhlik (eds) [1993] *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*, Vol. 2, East Lansing, MI: Philosophy of Science Association.
- Hume, David [1748] *An Enquiry Concerning Human Understanding*. Cited from Hume, *An Enquiry Concerning Human Understanding*, Charles W. Hendel (ed.), New York: MacMillan, 1955.
- Irvine, A.D. [1998] "Let Truth and Falsehood Grapple", *University of Toronto Quarterly* 67, 549-566.
- Irvine, A.D. [1997] "The Origins of Analytic Philosophy", *Russell: The Journal of Bertrand Russell Archives* 17, 79-87.
- Irvine, A.D. [1993] "How Braess' Paradox Solves Newcomb's Problem", *International Studies in the Philosophy of Science* 7, 141-60.

- Irvine, A.D. [1992] "Gluts, Gaps and Paradox", *Canadian Journal of Philosophy*, Supp. Vol. 18, 273-299.
- Irvine, A.D. [1991] "Thought Experiments in Scientific Reasoning", in Tamara Horowitz and Gerald J. Massey (eds), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 149-165.
- Irvine, A.D. [1989] "Epistemic Logicism and Russell's Regressive Method", *Philosophical Studies* 55, 303-327.
- Jackson, M.W. [1992] "The *Gedankenexperiment* Method of Ethics", *Journal of Value Enquiry* 26, 525-535.
- Janis, Allen I. [1991] "Can Thought Experiments Fail?", in Tamara Horowitz and Gerald J. Massey (eds), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 113-118.
- Johnston, Mark [1987] "Human Beings", *Journal of Philosophy* 84, 59-83.
- Kagan, Shelly [1988] "The Additive Fallacy", *Ethics* 99, 5-31.
- Kelly, Kevin [1994] *Out of Control*, Reading, MA: Addison Wesley.
- Kripke, Saul [1975] "Outline of a Theory of Truth", *Journal of Philosophy* 72, 690-716.
- Kitcher, Philip [1981] "Explanatory Unification". Cited from Boyd, Gasper and Trout (eds), *The Philosophy of Science*, Cambridge: MIT Press, 1991, 329-348.
- Kuhn, Thomas S. [1977] *The Essential Tension*, Chicago: The University of Chicago Press.
- Kuhn, Thomas S. [1964] "A Function for Thought Experiments", *L'aventure de la science, Melanges Alexandre Koyre*, Vol. 2, Paris: Hermann, 307-34. Cited from Kuhn, *The Essential Tension*, Chicago: University of Chicago Press, 1977, 240-265.
- Kujundzic, Nebojsa [1993] "How Does the Laboratory of the Mind Work?", *Dialogue* 32, 573-78.
- Laymon, Ronald [1991] "Thought Experiments of Stevin, Mach and Gouy: Thought Experiments as Ideal Limits and as Semantic Domains", in Tamara Horowitz and Gerald J. Massey (eds), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 167-191.
- Lennox, James G. [1991] "Darwinian Thought Experiments: A Function for Just-So Stories", in Tamara Horowitz and Gerald J. Massey (eds), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 223-245.
- Mach, E. [1883] *The Science of Mechanics*, J. McCormack (trans.), sixth edition, La Salle, Illinois: Open Court, 1960.
- Madell, Geoffrey [1985], "Derek Pafit and Greta Garbo", *Analysis* 45, 105-109.
- Martin, Raymond [1988] "Identity's Crisis", *Philosophical Studies* 53, 295-307.

- Massey, Barbara [1991] "Do All Rational Folk Reason As We Do? Frege's Thought Experiment Reconsidered", in Tamara Horowitz and Gerald J. Massey (eds), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 99-110.
- Massey, Gerald J. and Tamara Horowitz (eds) [1991] *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield.
- Massey, Gerald J. [1991] "Backdoor Analyticity", in Tamara Horowitz and Gerald J. Massey (eds), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 285-296.
- Maxwell, J.C. [1872] *Theory of Heat*, New York: D. Appleton and Co, 1972.
- Mill, John Stuart [1843] *A System of Logic, Ratiocinative and Inductive*, J.M. Robson (ed.), Toronto: University of Toronto Press, Routledge and Kegan Paul, 1973.
- Mohanty, J.N. [1991] "Method of Imaginative Variation in Phenomenology", in Tamara Horowitz and Gerald J. Massey (eds), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 261-272.
- Moor, James H. [1995] "Is Ethics Computable?", *Metaphilosophy* 26, 1-21.
- Moody-Adams, Michele M. [1990] "On the Alleged Methodological Infirmity of Ethics", *American Philosophical Quarterly* 27, 225-235.
- Nersessian, Nancy J. [1993] "In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling", in David Hull, Micky Forbes and Kathleen Okruhlik (eds), *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*, Vol. 2, East Lansing, MI: Philosophy of Science Association, 291-301.
- Norton, John D. [1996] "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy* 26, 333-366.
- Norton, John D. [1993] "Seeing the Laws of Nature", *Metascience*, n.s. 3, 33-38.
- Norton, John D. [1991] "Thought Experiments in Einstein's Work", in Tamara Horowitz and Gerald J. Massey (eds), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 129-147.
- Nowell-Smith, P.H. [1954] *Ethics*, Baltimore: Penguin Books.
- Nussbaum, Martha Craven [1989] "'Finely Aware and Richly Responsible': Literature and the Moral Imagination", in Stanley G. Clarke and Evan Simpson (eds), *Anti-Theory in Ethics and Moral Conservatism*, Albany: State University of New York Press, 111-34.
- Nussbaum, Martha Craven [1983a] "Flawed Crystals: James's *The Golden Bowl* and Literature as Moral Philosophy", *New Literary History* 15, 25-50.
- Nussbaum, Martha Craven [1983b] "Reply to Richard Wolheim, Patrick Gardiner and Hilary Putnam", *New Literary History* 15, 201-8.

- Okruhlik, Kathleen, Hull, David, and Micky Forbes (eds) [1993] *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*, Vol. 2, East Lansing, MI: Philosophy of Science Association.
- Oppenheimer, Joe and Norman Frolich [1997] "A Role for Structured Observation in Ethics", *Social Justice Research* 10, 1-21.
- Pascal, Louis [1980] "On Goodness, Badness and Indifference", *Inquiry* 23, 242-51. Cited from the abridged and reprinted paper "Judgement Day", in Peter Singer (ed.) *Applied Ethics*, Oxford: Oxford University Press, 1986, 105-123.
- Passell, Peter [1993] "Pac-Man It's Not: The Future of a Presidency and U.S. Health Care, Maybe", *The New York Times*, December 24, 1993, A16.
- Perry, John (ed.) [1975] *Personal Identity*, Berkeley: University of California Press.
- Pinch, Trevor, D.C. Gooding and Simon Schaffer (eds) [1989] *The Uses of Experiment*, Cambridge: Cambridge University Press.
- Plato, *Plato: The Republic*, Paul Shorey (trans.), Cambridge, Mass: Loeb Classical Library, 1930. Cited from Huntington Cairns and Edith Hamilton (eds), *The Collected Dialogues of Plato*, Princeton: Princeton University Press, 1989, 575-844.
- Plutarch "Life of Theseus", *Lives*, John Langhorne and William Langhorne (trans.), London: Printed for Edward and Charles Dilly, in the Poultry, 1770.
- Popper, Karl [1959] *The Logic of Scientific Discovery*, London: Hutchinson and Co.
- Putnam, Hilary [1983] "Taking Rules Seriously — A Response to Martha Nussbaum", *New Literary History* 15, 193-200.
- Putnam, Hilary [1973] "Meaning and Reference", *Journal of Philosophy* 70, 699-711. Cited from A.P. Martinich (ed.) [1985] *The Philosophy of Language*, second edition, Oxford: Oxford University Press, 1985, 308-315.
- Quine, W.V. [1990] *Pursuit of Truth*, Cambridge: Harvard University Press.
- Quine, W.V. and J.S. Ullian [1970] *The Web of Belief*, second edition, New York: Random House, 1978.
- Quine, W.V. [1969] "Epistemology Naturalized", *Ontological Relativity and Other Essays*, New York: Columbia University Press, 69-90.
- Quine, W.V. [1951] "Two Dogmas of Empiricism", *Philosophical Review* 60, 20-43. Cited from Quine, *From a Logical Point of View*, Cambridge: Harvard University Press, 1964, 20-46.
- Quinn, Warren [1993] *Morality and Action*, Cambridge: Cambridge University Press.
- Rawls, J. [1971] *A Theory of Justice*, Cambridge: Harvard University Press.
- Rawls, J. [1957] "Outline of a Decision Procedure for Ethics", *Philosophical Review* 66, 177-97.

- Reichenbach, Hans [1938] *Experience and Prediction*, Chicago: University of Chicago Press.
- Reid, Thomas [1785] *Essays on the Intellectual Powers of Man*, A.D. Woosley (ed.), London: MacMillan and Co., 1941.
- Resnick, Mitchel [1994] *Turtles, Termites and Traffic Jams*, Cambridge: MIT Press.
- Reynolds, Craig W. [1987] "Flocks, Herds and Schools: A Distributed Behavioral Model", *Computer Graphics* 21, Proceedings of SIGGRAPH '87, ed. Maureen C. Sttne, 25-34.
- Russell, Bertrand [1912] *The Problems of Philosophy*, Oxford: Oxford University Press, 1959.
- Russell, Bertrand [1907] "The Regressive Method of Discovering the Premises of Mathematics". Cited from Russell, *Essays in Analysis*, Douglas Lackey, (ed.), London: George Allen and Unwin, 1973, 272-283.
- Schaffer, Simon, Gooding, D.C., and Trevor Pinch (eds) [1989] *The Uses of Experiment*, Cambridge: Cambridge University Press.
- Schellenberg, Kathryn (ed.) [1994] *Computers in Society*, Guilford: Dushkin Publishing.
- Schelling, T.C. [1978] *Micromotives and Macrobehaviour*, New York: Norton.
- Schlipp, P.A. (ed.) [1959] *Albert Einstein Philosopher-Scientist*, La Salle: Open Court.
- Shoemaker, Sydney [1992] "Unger's Psychological Continuity Theory", *Philosophy and Phenomenological Research* 52, 139-43.
- Shoemaker, Sydney and Richard Swinburne [1984] *Personal Identity*, Oxford: Basil Blackwell.
- SimHealth [1993] *SimHealth: A Democracy and Society Computer Simulation, User's Manual*, Orinda: Sim-Business.
- Singer, Peter (ed.) [1986] *Applied Ethics*, Oxford: Oxford University Press.
- Singer, Peter [1974] "Sidgwick and Reflective Equilibrium", *Monist* 58, 490-517.
- Singer, Peter [1972] "Famine, Affluence and Morality", *Philosophy and Public Affairs* 1, 229-243. Cited from *Moral Philosophy: Selected Readings*, George Sher (ed.), San Diego: Harcourt Brace Jovanovich, 1987, 595-605.
- Skyrms, Brian [1996] *Evolution of the Social Contract*, Cambridge: Cambridge University Press.
- Skyrms, Brian [1966] *Choice and Chance: An Introduction to Inductive Logic*, third edition, Belmont, California: Wadsworth Publishing Company, 1986.
- Sorensen, Roy [forthcoming] "Review of Soren Haggqvist's *Thought Experiments in Philosophy*", *Theoria*.

- Sorensen, Roy [1995a] "Precis of Thought Experiments", *Informal Logic* 17, 385-387.
- Sorensen, Roy [1995b] "Sorensen's Reply to Bunzl and Feldman", *Informal Logic* 17, 399-405.
- Sorensen, Roy [1992a] *Thought Experiments*, Oxford: Oxford University Press.
- Sorensen, Roy [1992b] "Thought Experiments and the Epistemology of Laws", *Canadian Journal of Philosophy* 22, 15-44.
- Sorensen, Roy [1991] "Moral Dilemmas, Thought Experiments and Concept Vagueness", *Philosophical Studies* 63, 291-308.
- Straughan, R.R. [1975] "Hypothetical Moral Situations", *Journal of Moral Education* 4, 183-189.
- Swenson, Loyd S. [1972] *The Etherial Aether: A History of the Michelson-Morley Aether-drift Experiments, 1880-1930*, Austin: University of Texas Press.
- Swinburne, Richard [1992] "Discussion of Peter Unger's *Identity, Consciousness and Value*", *Philosophy and Phenomenological Research* 52, 149-152.
- Swinburne, Richard and Sydney Shoemaker [1984] *Personal Identity*, Oxford: Basil Blackwell.
- Thomson, Judith Jarvis [1986] *Rights, Restitution and Risk*, Cambridge: Harvard University Press.
- Thomson, Judith Jarvis [1971] "A Defense of Abortion", *Philosophy and Public Affairs* 1, 47-68. Cited from Peter Singer (ed.) [1986] *Applied Ethics*, Oxford: Oxford University Press, 37-56.
- Thornton, Stephen [1997] "Popper, Karl", *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/popper,karl/>, first posted: November 13, 1997. Cited from last modification: November 18, 1997.
- Toulmin, Stephen [1991] "How Important is Theory to Moral Inquiry?", *Medical Humanities Review* 5, 19-23.
- Ullian, J.S. and W.V. Quine [1970] *The Web of Belief*, second edition, New York: Random House, 1978.
- Unger, Peter [1996] *Living High and Letting Die*, Oxford: Oxford University Press.
- Unger, Peter [1992a] "Precis of Identity, Consciousness and Value", *Philosophy and Phenomenological Research* 52, 134-7.
- Unger, Peter [1992b] "Reply to Reviewers", *Philosophy and Phenomenological Research* 52, 159-76.
- Unger, Peter [1990] *Identity, Consciousness and Value*, Oxford: Oxford University Press.

- Unger, Peter [1982] "Toward a Psychology of Common Sense", *American Philosophical Quarterly* 19, 117-29.
- Wiggins, David [1967] *Identity and Spatio-temporal Continuity*, Oxford: Basil Blackwell.
- Whitbeck, Caroline [1996] "Ethics as Design: Doing Justice to Moral Problems", *Hastings Center Report* 26, 9-16.
- White, Stephen L. [1992] "The Desire to Survive", *Philosophy and Phenomenological Research* 52, 153-58.
- Wilkes, Kathleen V. [1988] *Real People: Personal Identity without Thought Experiments*, Oxford: Clarendon Press.
- Williams, Bernard [1970] "The Self and the Future", *Philosophical Review* 79, 161-180. Cited from John Perry (ed.) [1975] *Personal Identity*, Berkeley: University of California Press, 179-198.
- Wilson, R.W. [1979] "The Cosmic Microwave Background Radiation", in *Les Prix Nobel: 1978*, Stockholm: Almqvist and Wiksell International.
- Winkler, Earl R. [1996] "Moral Philosophy and Bioethics: Contextualism vs the Paradigm Theory", in L.W. Sumner and Joseph Boyle (eds), *Philosophical Perspectives on Bioethics*, Toronto: University of Toronto Press, 50-78.
- Winkler, Earl R. [1993] "From Kantianism to Contextualism: The Rise and Fall of Paradigm Theory in Bioethics", in Earl R. Winkler and Jerrold R. Coombs, (eds), *Applied Ethics: A Reader*, Cambridge: Blackwell, 343-365.
- Winkler, Earl R. [1991] "Against Ethical Theory", *Medical Humanities Review* 5, 25-28.
- Winkler, Earl R. and Jerrold R. Coombs (eds) [1993] *Applied Ethics: A Reader*, Cambridge: Blackwell.
- Wittgenstien, Ludwig [1958a] *The Blue and Brown Books*, Oxford: Basil Blackwell.
- Wittgenstien, Ludwig [1958b] *Philosophical Investigations*, Oxford: Basil Blackwell.
- Wollheim, Richard [1983] "Flawed Crystals: James's *The Golden Bowl* and the Plausibility of Literature as Moral Philosophy", *New Literary History* 15, 185-191.

Index of Thought Experiments & Related Examples

- Ayer, A.J.
 programmed automata, 185
- Brown, James Robert
 sum of an ascending series of positive integers, 33
- Club of Rome
 The Limits to Growth, 151, 153
- Danielson, Peter
 rationality and morality, 141, 150, 154
- Darwin, Charles
 wolf selection, 62, 150
- Darwin, George
 tulips and trumpets, 13
- Dennett, Daniel
 free will thought experiments, 185
 neurologist, philosopher, and oracle, 194
- Einstein, Albert
 clock in a box, 164
 light beam, 138
- Euclid
 proof of the existence of an infinity of primes, 44
- flight simulator, 154
- Galileo, Galilei
 falling bodies, 4, 28, 37, 39, 43, 44, 158
- Gillies, Donald/Adams and Leverrier
 anomalous orbits of Uranus and Mercury, 58
- Hacking, Ian
 microscopy, 199
 steam engine, 119
- Harris, John
 survival lottery, 69
- Hofstadter, Douglas
 sphex and free-will, 53
- Hon, Giora
 canals of Mars, 161
 N rays, 162
- Horowitz, Tamara
 Quinn's rescue dilemmas, 196
- Kagan, Shelley
 self-defense and the aggressor, 52
- Maxwell, James
 Maxwell's demon, 73
- Michelson-Morley
 speed of light, 15
- Nussbaum, Martha
 James' *The Golden Bowl*, 189
- Pascal, Louis
 evil king and eleventh man, 72
- Penzias and Wilson
 background radiation, 17
- Plato
 Gyges' Ring, 18
- Plutarch/Hobbes
 Ship of Theseus, 132, 195

Prisoner's Dilemma, 142

Putnam, Hilary

twin earth, 127

Rawls, John

original position, 70, 116

Reid, Thomas

brave officer, 137

Resnick, Michael

cocktail party grouping, 65

Reynolds, Craig

flocking behaviour, 67, 140, 153

Schelling, Thomas

segregated neighbourhoods, 65

Schrödinger, Erwin

Schrödinger's cat, 75

Searle, John

Chinese room, 27

Shoemaker and Unger

Brainland, 186, 195, 200

SimHealth, 148, 152

Sorensen, Roy

bizarre counterfactuals, 98

universe doubled, 97

vacuist and plenist debate, 175

Stevin

Stevin's chain, 39, 43, 100, 139

Thomson, Judith Jarvis

ill violinist, 8, 27, 43, 45, 49, 100, 158, 191

people-seeds, 119

surgeon and five patients, 21, 43, 69, 97, 100

Unger, Peter

AFGP tests, 172

shallow pond and envelope, 50, 199

Whitbeck, Caroline

bridge design, 118

Wiggins, David

brain fission, 133

Wilkes, Kathleen

amœbae, 95, 97

Williams, Bernard

switching A and B, 193

Winkler, Earl

cardiac case, 60

MS case, 107, 114