SCIENCE FICTION THOUGHT EXPERIMENTS IN BIOETHICS

by

Jasmine Smart

Submitted in partial fulfilment of the requirements
for the degree of Master of Arts

at

Dalhousie University
Halifax, Nova Scotia
December 2012

DALHOUSIE UNIVERSITY

DEPARTMENT OF PHILOSOPHY

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "SCIENCE FICTION THOUGHT EXPERIMENTS IN BIOETHICS" by Jasmine Smart in partial fulfilment of the requirements for the degree of Master of Arts.

Dated: December 12, 2012

Supervisor: _____

Readers: _____

_____

DALHOUSIE UNIVERSITY

DATE: December 12, 2012

AUTHOR: Jasmine Smart

TITLE: Science Fiction Thought Experiments in Bioethics

DEPARTMENT OR SCHOOL: Department of Philosophy

DEGREE: MA CONVOCATION: May YEAR: 2013

_____

Signature of Author

Table of Contents

Abstract

Science fiction is particularly apt as bioethical thought experiment. In considering the theories of James R. Brown, John D. Norton and Marco Buzzoni, I suggest that mental-modeling theories afford the best explanation for what thought experiments can do. I propose a version of mental modeling that has the flexible modalities of experience found in Nancy J. Nersessian's account, combined with Nenad Miščević's compelling vision of how existing knowledge is used to create mental models, and Tamar Gendler's use of schemas to understand ethical thought experiments.

Bioethics makes use of thought experiments' capacity to move from abstraction to discrete instances. Sometimes thought experiments will be better, and sometimes real cases will be unavailable. Given the cognitive advantages that access to mental models provides, thought experiments will be of use in the field of bioethics.

To identify literature that is thought-experimental I look to Geordie McComb's family resemblance theory, and consider accounts of literary thought experiments by Noel Carroll and Edward Davenport. Extended narratives will in some cases be more useful for ethical understanding than philosophical thought experiments. Science fiction has this same advantage: as ethical narrative it is detailed and humanized. In addition the speculative nature of science fiction lends itself to the exploration of new and emerging sciences and technologies including those in the field of bioethics.

Acknowledgements

First and foremost I wish to thank Letitia Meynell, my supervisor, without whom I would have been lost. Her endless support and expertise have been invaluable.

This dissertation would not have been possible without the guidance and the help of my committee, Kirstin Borgerson and Mike Hymers. It is with immense gratitude that I acknowledge the insight and clarity provided by Kirstin Borgerson. I am indebted to Mike Hymers for his precision and thoroughness.

Chapter 1 Introduction

What Are Thought Experiments?

I'm interested in looking at thought experiments as methods of imagining alternative possibilities. In the introduction to the Stanford Encyclopedia of Philosophy entry on the topic, James R. Brown and Yiftach Fehige describe thought experiments as "devices of the imagination used to investigate the nature of things",[1] and identify some important features of thought experiments including that thought experiments are generally narratives, are used in a variety of disciplines including but not limited to philosophy, implicate the imagination, and are experiential. Not all counterfactuals are thought experiments, as a counterfactual may have no imaginative, experiential, perceptual, or experimental qualities. However, all thought experiments are counterfactual, otherwise they would be case studies or real experiments.[2] It is common to find thought experiments that begin by saying 'imagine yourself...', or 'picture yourself...'. A thought experiment often asks that you imagine, visualize, put yourself in the place of, experience the unfolding of events, or traverse a causal chain. What distinguishes thought experiments from fiction simpliciter is that thought experiments have a cognitive upshot which related to the real world and not only the fictional worlds they inhabit.

Project Description

The first half of my thesis is divided into four chapters on thought experiments in

---

[1] Brown and Fehige n.p.

[2] Thanks to my supervisor, Letitia Meynell, for raising this issue.

science, mental modeling, thought experiments in ethics, and finally bioethics. My focus is on ethical and scientific thought experiments because my subject is the use of thought experiments in bioethics. There are fascinating thought experiments in other fields, notably epistemology and personal identity; however, I have focused on those disciplines that are most closely related to bioethics. Scientific thought experiments produce knowledge, and ethical thought experiments unearth our intuitions and explain or motivate ethical actions. How thought experiments can produce knowledge, and indeed whether they can produce knowledge, is a major question in the philosophy of science. I will introduce two thought experiments by Galileo to show how thought experiments are like real experiments, and then look at two theories, those of Brown and Buzzoni, that relate thought experiments and real experiments. Finally, I suggest that mental modeling theories of thought experiments provide an answer to this question.

Ethics is another rich vein of thought experimentation. Ethical thought experiments have the disadvantage of describing the realm of morality, which is less widely agreed upon than the physical world, and seemingly less amenable to definitive proofs. In describing ethical thought experiments as 'intuition pumps', the issue of moral intuitions is problematized, though of course moral intuitions are not the only intuitions that are elicited by thought experiments. That thought experiments draw out moral and conceptual intuitions is not in serious contention—the more interesting question is whether or not the intuitions unearthed in this manner are good ones. This issue is connected to the larger question of the status of moral intuitions. I will argue that moral intuitions are a significant problem only for justificatory thought experiments, and not for

explanatory or motivational ones. I will look at trolley problems as explanatory thought experiments, and Singer's drowning child as a motivational thought experiment. Thought experiments in science, mental modeling, and thought experiments in ethics are brought together in the concluding section on bioethics. The overall conclusion for the first half of my thesis is that thought experiments can be useful in the field of bioethics. If we give up thought experiments, then we lose elements of ethics and science that are valuable to bioethics.

In the second half I argue that science fiction thought experiments not only can be bioethical thought experiments, but that science fiction thought experiments may indeed be better because as ethical narratives they are more detailed and humanized, and the speculative nature of science fiction lends itself to the exploration of new and emerging sciences and technologies.[3] I will argue that fiction can act as, and be, thought experiments. I will begin by considering objections to the idea of literary thought experiments. For the role of ethics in literary thought experiments I will look at Carroll's article about virtue ethics in literature. However, while narratives act as tools of ethical analysis, it is not only ethics that is enacted in literature. Davenport's article takes scientific thought experiments as a starting point. Finally, I will examine a family resemblance theory by Geordie McComb that provides the means to identify thought experiments in fiction.

Having established the legitimacy of literary thought experiments, the final chapter

---

3 We might also ask if there are questions relevant to bioethicists that go beyond a focus on new technologies. “Bioethics has been critiqued for being too focused on the ethics of new technologies, and my emphasis on science fiction might exacerbate this.” (Thanks to my second reader, Kirstin Borgerson, for raising this issue)

looks at science fiction. In the chapter on science fiction I begin by establishing a working definition of science fiction for the purposes of this thesis that is compatible with literary thought experiments and includes the elements of science, imagination and speculation. The use of alternative worlds, and the vivid imagining of what could be supports the use of science fiction thought experiments, and fits with mental modeling theories of thought experiments. Mental modeling accounts are exemplified in science fiction thought experiments because mental modeling emphasizes the role of imaginative capabilities and of using imagination to grasp possibilities. I will include a discussion of utopias because they share characteristics and overlap with science fiction, and because the particularly moral characteristic of utopias is useful in justifying why science fiction thought experiments are particularly relevant to bioethics. The questions I consider are what science fiction is, how it connects to mental modeling and to imagining possible worlds, and finally why science fiction is a good fit for bioethics. To answer this final question I return to my definition of science fiction, which claims that science fiction has to be about science, and I look to utopias and argue that they are related to science fiction and that they, and literature generally, have narrative characteristics that enhance ethical understanding. Science fiction is thus a good fit for bioethics because it brings together science and ethics.

Why Write About Thought Experiments?

I will offer a foil to motivate my discussion of science fiction thought experiments in bioethics. In “The Trouble with Thought Experiments", Jeremy Goodenough decries the use of thought experiments in general, and specifically in bioethics. Goodenough argues

that as a means of accessing moral and conceptual intuitions thought experiments prove problematic, and that part of the problem is that, "Trying to define what a thought experiment is has proved to be a difficult task."[4] I agree that thought experiments are hard to define; however, it is not uncommon to make use of things that we cannot precisely or uncontroversially define, and that something is merely difficult to define is not a strong charge against it.[5] And so I do not offer a definition of thought experiments here, but instead provide an account of what they do.

Goodenough also questions the use of examples at all, asking, "But are real cases any better? They certainly test our concepts, showing that in extraordinary cases our everyday conceptual framework struggles to accommodate the case. And in the case of genuine medical dilemmas, they tell us something, even if it is only that life occasionally throws up insoluble difficulties. What they don't seem to do is provide anything more positive than this."[6] Having grudgingly allowed some place for real cases, Goodenough claims that thought experiments are always worse than real examples and that 'far-fetched' or 'science-fiction' thought experiments are particularly problematic. What Goodenough means by science fiction thought experiments are those that are far-fetched or include "some kind of impossibility."[7] My argument is that thought experiments are not only as useful as real examples; they can sometimes be better. Moreover, being far-fetched is not

---

4Goodenough p.7

5 In the Proceedings of the XXth World Congress of Philosophy the is a paper by Daniel Andler titled "The Undefinability of Analytic Philosophy" in which Andler writes that analytic philosophy itself is undefinable; "Whereas the existence of analytic philosophers is uncontentious, and their identification easy enough in most cases, it is much harder to say what analytic philosophy is, even relative to a given conception of philosophy *tout court."* Should we draw from this paper the conviction that analytic philosophy is hard to define, that alone does not seem sufficient reason to give up the enterprise. Arguing from analogy, that thought experiments are hard to define does not mean we should give them up.

6Goodenough p. 11

7Goodenough p. 7

necessarily a problem, and indeed I will show that some thought experiments taken from science fiction literature are particularly well suited to use in bioethics because they bring together elements of science and ethics.

Thought experiments are useful in contexts where we can't use real experiments, and even in situations where real cases are possible, thought experiments may be better than any real case. I will defend the mental modeling approach to thought experiments and show what can be done using mental modeling. One of the strengths of this analysis is that by creating models instead of using propositional reasoning there are problems that are more easily solved. Indeed, there may be problems that we can solve using models that we cannot solve using propositional reasoning alone. I will also argue that the flexibility to create first-person narratives can be an important tool in identifying or eliminating bias by discussing the use of the first person in Thomson's violinist thought experiment. I will make a case for thought experiments in extended narratives creating a more vivid and complete picture than is possible thorough case studies or argument in the section on utopias and dystopias. In calling for caution in the use of thought experiments Goodenough risks losing these important functions of thought experimentation.

Goodenough claims that there are sufficient real situations to make hypotheticals superfluous. However, not only can thought experiments sometimes do things that real experiments cannot, in bioethics there will be situations for which we do not yet have instantiations; there is a place for thought experiments to fill these gaps in domains where there have not yet been applications (e.g. new technologies, approaches, procedures). In addition there will be experiments that would be unethical to actually perform, and in

these cases thought experiments can take the place of the real cases that are unavailable to us.

Science fiction thought experiments fall squarely into Goodenough's category of far-fetched thought experiments. In response to claims that Goodenough makes that 'far-fetched' thought experiments are particularly problematic, I will make a case first that some works of science fiction are thought experiments, and that moreover such instances can be useful. One of the reasons Goodenough gives to be wary of thought experiments is the lack of context in such hypotheticals. I suggest that literary forms do not suffer from this criticism, as they are often as rich in detail as case studies if not richer. There is a great deal more background available in a novel than in case notes. Literary thought experiments provide rich background and contextual information which is one reason to think that science fiction thought experiments may be apt for bioethics.

I will argue that thought experiments from science fiction are particularly apt for bioethics. In looking at bioethics as a commingling of scientific and ethical enquiries, an account of thought experiments is required that includes both. Using mental modeling I will argue that thought experiments are cognitive tools of startling power and precision that enable problem-solving by means not available through argumentation alone. Not only are thought experiments pervasive and important in both the philosophy of science and ethics, they play roles in the production of knowledge and conceptual analysis that cannot be replaced either by actual experimentation, argument or by case studies. Mental-modeling accounts of thought experiments offer explanations for both scientific and ethical thought experiments that centre on the human ability to imagine possibilities.

This imaginative ability is exemplified in the use of science fiction thought experiments such as those that consider plausible future technology, the potential moral consequences of a particular use, its widespread adoption, and the result of any of these courses on society or individuals. Once I have established that thought experiments do have value in bioethics, the question is then twofold: when does science fiction properly function as a thought experiment, and when is this useful in bioethics. Bioethics falls in fertile ground between the medical and biological sciences and ethics, and science fiction is particularly apt in such instances, as it often deals with the applications of technology and the moral consequences thereof.

Goodenough warns against such uses, writing that

> There is an old lawyers' saying: "hard cases make for bad law." My own suspicion is that hypothetical cases often make for worse, and impossible cases are the worst of all. The more they incorporate impossibilities, the more problems they face. And even where we use actual cases, their degree of improbability limits their use in helping us to develop problem-solving abilities. There is, then, nothing wrong with using hypothetical cases: but the more hypothetical they are, the more they should carry a 'handle with care' sticker.[8]

"Be careful" is generally good advice, but saying 'be careful on the bike ride home' does not mean that it is a bad idea to go home, or to ride your bike. If we were to give up on every realm of enquiry that includes the possibility of misleading results, we would be left with nothing. If we dismiss thought experiments, we are left impoverished. I propose a robust conception of thought experiments, based on mental modeling. By doing so I hope to show, not what thought experiments are for, but how they can create knowledge.

8Goodenough p. 12

Chapter 2 Thought Experiments in Science

Introduction

Thought experiments produce knowledge. In this chapter I will suggest that they do this by being like real experiments and so shape the direction of scientific progress. To establish this I will present two thought experiments by Galileo, which I will then examine using the theories of James R. Brown, John D. Norton and, briefly, Marco Buzzoni. I will conclude that these theorists all capture important facets of what thought experiments can do. Buzzoni shows that thought experiments are indeed like real experiments, Norton captures the role of contradictory thought experiments and the epistemic value of identifying a good thought experiment, and Brown gives a very useful taxonomy of the epistemic functions that thought experiments have.

In the introduction to the book *Thought Experiments in Science and Philosophy* Tamara Horowitz and Gerald J. Massey reflect on the thought experiment as a method both of science and philosophy;

> The line between science and philosophy is sometimes drawn at observation. Observation itself can be passive and even unplanned, or active and artfully contrived. Observation of this second, deliberate sort is commonly know as *experimentation*. But when one reflects that scientific experiments are at least as likely to be *thought experiments* (*Gedankenexperimente*) as real ones, even the seemingly hard-headed appeal to observation to demarcate science from philosophy begins to look fanciful. Why? Because philosophers conduct thought experiments, too.[9][10]

9Horowitz and Massey p.1

10 Thanks to my third reader, Michael Hymers, for pointing out that this quote presupposes that thought experiments *are* experiments of a special sort. And for the following quote from Wittgenstein offering an opposing view: "What Mach calls a thought experiment is of course not an experiment at all. At bottom it is a grammatical investigation" (Wittgenstein, *Philosophical Remarks*, s1).

In this section I explain what scientific thought experiments are and what they can do, preparatory to the next chapter in which I will argue that mental modeling is the best way to understand thought experiment. In explaining what scientific thought experiments are, I hope to distinguish them from ethical thought experiments in order to show how bioethical thought experiments contain elements of both. I begin with scientific thought experiments because they have received more scholarly attention, and most clearly demonstrate the creation of new knowledge without infusions of new empirical data. This is also the purpose of the two example thought experiments from Galileo which I have included, both to show that thought experiments are like real experiments, and to show that thought experiments produce knowledge.

Galileo

> Without experiment, I am sure that the effect will happen as I tell you,
>
> because it must happen that way -Galileo, *Diologo*[11]

I begin by introducing two of Galileo's thought experiments to show how they are similar to real experiments. Galileo's Salviati ship thought experiment counters an Aristotelian thought experiment. Aristotle's tower attempted to show that it is impossible that the Earth is in motion, on the grounds that if the Earth were moving objects would not fall straight down; they would always fall down and behind in the direction from which the earth had come. If you were to stand at the top of a moving tower, then a dropped object would not fall at the base of the tower, but a distance from the tower in the opposite direction from which the tower is moving—in its wake. Galileo describes

11As quoted in Brown (2004) p. 27

Aristotle's tower in the following way:

> Aristotle says, then, that a most certain proof of the earth's being motionless is that things projected perpendicularly upward are seen to return by the same line to the same place from which they were thrown, even though the movement is extremely high. This, he argues, could not happen if the earth moved, since in the time during which the projectile is moving upward and then downward it is separated from the earth, and the place from which the projectile began its motion would go a long way toward the east, thanks to the revolving of the earth, and the falling projectile would strike the earth that distance away from the place in question.[12]

Having thus explained Aristotle's view, Galileo convinces us that a stone dropped from the mast of a moving ship would act in the same manner as a stone dropped from a tower on a moving earth. Having established that they are analogous, Galileo claims that what will occur in both cases is that the stone will fall to the base, and *not* in the wake of either the ship or the tower.

> For anyone who does [attempt this experiment] will find that the experiment shows exactly the opposite of what is written; that is, it will show that the stone always falls in the same place on the ship, whether the ship is standing still or moving with any speed you please. Therefore, the same cause holding good on the earth as on the ship, nothing can be inferred about the earth's motion or rest from the stone falling always perpendicularly to the foot of the tower.[13]

What is of particular interest is that this thought experiment works very well in our mind, and only with more difficulty and room for error in the external world. The requirement that the boat's motion be constant and smooth is much easier to imagine than to actualize, and thought experiments allow us to imagine that distracting factors like wind are not at issue, instead of real experiments which requires laborious controlling for their presence

---

12Galilei p. 143
13Galilei p. 146

or absence. Thus the thought experiment is better than the real experiment, and replaces it by making the actual experiment unnecessary because we can so clearly tell what will happen. This does not replace experiential evidence, it is this experience that tell us that objects will fall down rather than float or fall upwards. The thought experiment allows us to imagine what should happen using our experience.[14] By stipulating rules for representing the events Galileo creates circumstances that can be clearly imagined, giving a result in our minds that is clearer than the results that would be obtained if the experiment were actually performed. It convinces us like a real experiment, but in it's clarity it does so better.

Thought experiments are imagined rather than physically performed, and subsequently some have been vindicated with empirical support as strong as for any knowledge that we have. For instance, that the earth is indeed moving and circles the sun is well confirmed by empirical evidence. That the earth is, to the *very* best of our knowledge, moving is a reason to think that Galileo's Salviati Ship thought experiment was correct. Thus thought experiments stand alone, without a physical experiment to demonstrate the claims they advance, and I will argue that they produce knowledge that we later see is true by the best evidential standards. This is good reason to think that thought experiments are valuable and produce knowledge. There is the concern that it is actually the subsequent proof that does the epistemic work in such cases,[15] and so what is necessary is an account of thought experiments that accounts for the thought experiment predating the proof, and gives the thought experiment the appropriate epistemic weight,

---

14Thanks to my second reader, Kirstin Borgerson, for raising this objection.
15Thanks to my supervisor, Letitia Meynell, for raising this objection.

if the thought experiment itself is to be the site of knowledge production.

Galileo's falling bodies is also a thought experiment that demonstrated inconsistencies in Aristotelian physics. Galileo challenged the claim that heavier objects fall faster than lighter objects. If we were to imagine standing at the top of a tower holding a cannonball attached to a musketball, the cannonball would be the heavier object, and the musketball the lighter. In attaching them together a dilemma is created: will the cannonball and musketball, being heavier than the cannonball alone, fall faster than the cannonball, or will the mustketball, being lighter, slow the rate of descent of the heavier cannonball?[16] According to Aristotle's theory both should occur, and since the cannonball attached to the musketball cannot fall both faster and slower, a contradiction is produced.

In both Galileo's falling bodies and Salviati ship thought experiments there is both a negative and a positive claim. The negative claim is simply that Aristotle was wrong. Showing a contradiction in the falling bodies thought experiment exposed an inconsistency in certain claims of Aristotelian physics. That one thought experiment can discredit another thought experiment elicits questions about how knowledge is produced by thought experiments that will be considered later in this chapter.

Brown, Frappier and Meynell ask this very question about the falling bodies thought experiment writing: “In the Galileo case for instance, we arrive at the result without the benefit of new empirical evidence and without deriving the result from things we already know.”[17] This production of new knowledge is both the discrediting of the former claim, but also and more strikingly the positive claim of new knowledge: “The usual reading of

16Brown (1991) p.1
17Brown, Frappier and Meynell pp. 3

this thought experiment is that it not only signifies the end of Aristotle's theory; it is also the answer to how fast these objects fall—obviously they all fall at the same rate."[18] Thus we see how knowledge, in particular scientific knowledge which is not derived from existing knowledge claims, can be produced without new evidence.

Thought Experiments and Real Experiments

Marco Buzzoni has a theory of thought experiments in which only those thought experiments which could be enacted as real experiments are valid. In "Empirical Thought Experiments: A Transcendental-Operational View" he writes that "All thought experiments must be thought of as translatable into real ones, and all real experiments as realisations of thought ones."[19] He goes on to say that "What thought experiments have over and above real experiments is the mere fact that they exist in a purely hypothetical sphere; what real experiments have over and above thought experiments is the mere fact that they overstep the sphere of the possible, in the experiment's real execution."[20] I believe there are problems with this view, as some thought experiments cannot be cast in terms of real experiments that we have not yet run, or even as real experiments that we do not yet have the means of running. Einstein's beam of light[21] is surely one, as we are never going to be able to run alongside a beam of light to watch what it does. A very strong argument would be necessary for us to give up such thought experiments which cannot even in principle be performed. However what I would like to draw from Buzzoni is the idea that thought experiments and real experiments are genuinely similar. It may be

---

18 Brown, Frappier and Meynell p. 3
19 Buzzoni p.1 italics removed.
20 Buzzoni p.1 italics in original.
21 Einstein asks us to imagine ourselves running at the speed of light to observe the behaviour of a beam of light.

that we can understand the act of reading about an experiment performed by another as running the real experiment in our minds as though it were a thought experiment. Moreover, it may be that thought experiments like Einstein's beam of light that are never attempted are also like real experiments and have epistemic similarities to real experiments. Towards that project I will now look at Brown's a priori account of thought experiments, which are one explanation for how thought experiments work, given a very robust concept of what can be discovered a priori. I will then contrast Brown's Platonic theory with John Norton's theory of thought experiments as arguments.

The two best known accounts of thought experiments are those of Jim Brown and John Norton. In the following summary from Norton's University of Pittsburgh webpage Norton draws out the battle lines, characterizing the division as that between belief in Platonic forms and understanding thought experiments as arguments that are charmingly couched as word pictures:

> How do thought experiments give us knowledge of the world? Many suppose that there is some special sort of epistemic power inherent in thought experiments and that they may even open windows through which we can perceive the Platonic forms of the laws of nature themselves. In a series of studies, I have defended a deflationary account of the nature of thought experiments: they are merely picturesque arguments, I say, and have no special epistemic powers beyond those of ordinary argumentation.[22]

Brown's view places thought experiments in the realm of a priori knowledge. In this thesis I adopt a mental modeling approach, and in the next chapter I show how mental modeling can explain what thought experiments are capable of without recourse to a priori truth. In addition, mental modeling gives a plausible account of thought experiments that is non-propositional.

22http://www.pitt.edu/~jdnorton/homepage/research/thought_expt.html

The Platonic Realm

The poet's eye, in a fine frenzy rolling,
Doth glance from heaven to earth, from earth to heaven,
And as imagination bodies forth
The forms of things unknown, the port's pen
Turns them to shapes, and gives to airy nothing
A local habitation, and a name.
-Shakespeare, *A Midsummer Night's Dream*[23]

In *The Laboratory of the Mind,* James R. Brown suggests a taxonomy for thought experiments. He explains that thought experiments are like real experiments in having different functions: "...they 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."[24] Brown distinguishes between thought experiments that make positive and negative claims, calling these constructive and destructive thought experiments respectively.[25]

A destructive thought experiment shows a claim to be false, or works against a given theory.[26] Constructive thought experiments are further divided into direct, conjectural and mediative, all of which make a positive claim.[27] Mediative thought experiments work from already articulated theories and illustrate a problematic aspect, either to clarify or

---

23As quoted in Walton p.1
24Brown p.33
25Brown p.33
26Brown p.33
27Brown p.33

gain support for the controversial feature.[28] Conjectural thought experiments do not start from an established theory, but instead propose a thought-experimental result from which theory is developed; it is the thought-experimental result that is controversial and that calls for a theory to explain it.[29] Brown gives Newton's bucket as an exemplar of conjectural thought experiments. Newton's bucket is a thought experiment in which we imagine a universe empty except for a bucket partially filled with water, which is tied to a rope. This rope is twisted around, so that when the bucket is released the cord untwists, and the bucket rotates. When the bucket is released, the water is level and does not move relative to the bucket. Soon the water and bucket will move in respect to one another, though the surface of the water remains level. And then the bucket and water 'stop' and are no longer in motion in relation to the other, but the surface of the water is no longer flat; it is concave.[30] There is no relative motion to provide an explanation for the concavity of the water; therefore absolute space was proposed by Newton to explain the phenomenon.[31] Direct thought experiments are much like conjectural thought experiments, but the thought experimental result is unproblematic, and the issue is in taking the thought experimental results and moving to a theory which fits the results.[32] I will return to these categories of mediative, conjectural, and direct constructive thought experiments, and destructive thought experiments, when I will apply these categories to ethical thought experiments.

Brown has a distinct category for any thought experiment that is both destructive and

---

28Brown p.36
29Brown p.40
30Norton (2004) p.45
31Norton (2004) p.45
32Brown p.41

constructive, which he calls Platonic thought experiment and describes as " ...a priori in that it is not based on new empirical evidence nor is it merely logically derived from old data".[33] I do not propose to advocate an a priori or Platonic theory of thought experiments; my interest here is in particularly strong examples of thought experiments that produce knowledge without the injection of new empirical evidence. Brown uses Galileo's falling objects thought experiment to illustrate his category of Platonic thought experiments. Not only does Galileo make use of no new empirical data; Brown claims that it can be shown that the falling objects thought experiment is neither logically deduced from old data, a logical truth, or the simplest adjustment to Aristotle's theory of falling objects. That it is not a logical derivation from the existing data is thus explained by Brown:

> The premises of such an argument could include all the data that went into Aristotle's theory. From this Galileo derived a contradiction. ... But can we derive Galileo's theory that all bodies fall at the same rate from these same premisses? Well, in one sense, yes, since we can derive anything from a contradiction; but this hardly seems fair. What's more, whatever we can derive from these premises is immediately questionable since, on the basis of the contradiction, we now consider our belief in the premisses rightly to be undermined. Might Galileo's theory be true by logic alone? To see that the theory that all bodies fall at the same rate is not a logical truth, it suffices to note that bodies might fall with different speeds depending on their colours or on their chemical composition.[34]

Thus it appears that we gain knowledge via thought experiments that not only contradict existing claims, but also advance a new knowledge claim, and that this knowledge is more than just an adjustment of the existing theory to account for new data (of which there is none). This robust claim that thought experiments produce knowledge is what I

33Brown p.77
34Brown p.79

am most interested in claiming for scientific thought experiments; however, I will suggest that mental modeling theories of thought experiments are a better account for how this is possible in the following section on mental models. In the next chapter I will argue that mental modeling provides the most convincing account of thought experiments, and clarifies the function of thought experiments. In addition to those advantages, mental modeling does not entail the metaphysical challenges that accompany a Platonic theory of thought experiments such as that of Brown.

Thought Experiments and Scientific Progress

In this section I look at an account of scientific progress that emphasizes the role of imagination. This account treats imagination, and so on my account thought experiments, as leading the advance of human knowledge. These thought experiments are like real experiments in that they produce knowledge and direct change in theories. The ideas here will carry through in my insistence that science fiction literature can contribute to bioethics because it includes useful aspects of both science and ethics, and in the idea that we can learn about the world through thought experiments in fiction.

Edward Davenport argues for looking at thought experiments that stand alone in his article, "Literature as Thought Experiment (On Aiding and Abetting the Muse)". The understanding that Davenport proposes is based in the use of thought experiments in the sciences. Davenport gives a quote from Hanson (in *Observation and Explanation*) that positions thought experiments as potentially deserving more weight than empirical evidence: "The failure of experimental results to support anterior theoretical reflections--this has always been, for some, an initial indication of something wrong in the

experimental design itself. Herein lies the power of gedankenexperiments, such as Galileo's Pisa-cannonballs, Newton's bucket, Einstein's elevator, Schrodinger's cat, etc.; the theoretical issues in such examples just overwhelm the virtues of pushing or pulling or cutting or heating chunks of matter in order to show "what is the case" to the unconvinced'."[35] Davenport also quotes Albert Einstein and Leopold Infeld from their book, *The Revolution of Physics,* saying that "we have seen that this law of inertia cannot be derived directly from experiment, but only by speculative thinking consistent with observation. The idealized experiment can never be actually performed, although it leads to a profound understanding of real experiments."[36] This view proposes thought experiments that stand alone, and influence science. A similar view of stand-alone thought experiments that affect the progress of science comes from Arthur Stinner.

In "Scientific Method, Imagination, and the Teaching of Physics", Arthur Stinner uses thought experiments to reflect on the history of scientific thought about freefall as an illustration of the role of imagination in scientific progress. The point is that imagination is equal in importance to discovery in the progress of science. Stinner describes a naive view of scientific method in which regularities in nature are observed, and from these observations laws of nature are discovered: "There is still a wide-spread and pervasive belief that scientists use a specifiable and teachable method in going from observation to establishing laws and theories, namely the *scientific method*."[37] Stinner notes the contributions that Popper and Kuhn made to the understanding of scientific progress as a history of human understanding, mediated by the concepts and ways of understanding

35Davenport p. 282
36Davenport p. 282
37Stinner p. 335 italics in original

that scientist have available to them. Imagination as well as understanding is given a role in scientific practice by Kuhn as well, not so much in the practice of normal science as in moments of great conceptual change when it is up to the scientist to reinterpret what information is salient and what it means. Stinner writes that "There is, in a sense, the suggestion that the scientist struggles toward understanding the world, much like the artist strives to interpret it. After all, both are using imagination to see new patterns emerging from a web of constraints."[38] The idea that imagination is enriched or empowered by the use of thought experiments is illustrated in the changes to the modes of thought about freefall.

> An Aristotelian would see free fall as a natural motion that requires no other explanation or quantitative description. A Galilean would see free fall as a constantly accelerating motion, where both instantaneous velocity and acceleration are defined in terms of time and distance: it is a law-like motion but not natural in the Aristotelian sense. Natural motion now is understood as the unimpeded ("inertial") motion of an object circumnavigating the earth. A Newtonian would see free fall as the motion determined by the inverse square law of gravitational forces and the second law of motion. Natural motion for Newton now becomes a thought experiment and is pictured as the inertial motion of an object in deep space with zero net force and traveling at a constant speed in a straight line. For an Einsteinian, free fall is seen as motion in a four dimensional continuum of space and time. Natural motion now is seen as the motion of a free particle along a geodesic, the path of minimal separation.[39]

In considering Galileo's falling objects, I have already introduced one thought experiment about free fall in the first section of this chapter. Two equally famous thought experiments have been implicated in the development of Newton's and Einstein's views of freefall. Two spheres rotating in a void is a thought experiment created by Einstein in

---

38Stinner p. 336
39Stinner p. 337

which Newton's idea of inertial space was challenged:

> In 1916, Einstein formulated this worry in a thought experiment. He imagined two fluid bodies in a distant part of space. These bodies, the reader quickly infers, are like stars or planets, which form roughly spherical shapes under their own gravity. Einstein further imagines that there is relative rotation between the two bodies about the axis that joins them. This relative rotation is verifiable by observers on each body, who can trace out the motion of the other body. Each would judge the other to be rotating. It can happen in ordinary Newtonian physics that one of these bodies is not rotating with respect to an inertial frame and the other one is. In that case, the second rotating body will bulge. This effect arises on the earth. It rotates about the axis of its north and south poles. It bulges slightly at the equator as a result of centrifugal forces that seek to fling the matter of earth away from this axis. It would be entirely unacceptable, Einstein now asserted, were this to happen to two spheres in an otherwise empty space. For there is no difference in the observable relations between the two spheres. Each rotates with respect to the other. So why should just one bulge? Newton's absolute space or inertial systems, Einstein protested, was an inadequate explanation. Einstein demanded something observable to make the difference... In the case of Einstein's two fluid spheres, the bulge of one of them would now be explained by the fact that this bulging sphere was rotating with respect to all the other masses of the universe, whereas the other sphere was not. That would be the observable difference between the two fluid bodies. [40]

Thus we see that Einstein's spheres yields the opposite of the result of Newton's bucket. That opposite intuitions may be elicited is a problem for thought experiments in science as well as in ethics, and indeed for thought experiments more generally. Earlier in this chapter I described Galileo's Salviati Ship thought experiment, which was written in opposition to Aristotle's tower thought experiment, which was intended to convince of the opposite view—that the earth could not be moving around the sun.

This progressive chain of thought experiments appears to have shaped our

---

40Norton, <http://www.pitt.edu/~jdnorton/teaching/HPS_0410/chapters/general_relativity_pathway/index.html>.

understanding of free fall, and these changes came about in part because of counter-thought examples that created ways of thinking about the same problems. In the following section I will look at what it means to have thought experiments that express opposing views.

Thought Experiments as Arguments

John D. Norton suggested that thought experiment and anti-thought experiment pairs[41] are a test of viable epistemology for thought experiments.[42] An example of opposite intuitions evoked by thought experiments can be seen in a thought experiment pair proposed by Norton about whether or not space is infinite.[43] Aristotle claims that the universe cannot be infinite, because if the universe were infinite, smooth rotation would be impossible. Imagine a person pointing and that the line that she points along is infinite and rotating around a point. Now imagine a line parallel to the line of pointing. As the person rotates, the lines which begin as parallel must intersect if the second line is stationary. But it seems that we cannot find the first point where the lines will intersect. They are parallel at the starting point, and so do not intersect, but if the universe is infinite, then for any angle other than 0 there must be a smaller angle at which the two lines intersect. And yet they do not intersect at 0 degrees, thus it seems the pointing line never intersects the parallel line, which is impossible if the pointer is rotating, giving reason to believe that the concept of infinite space is impossible.[44] An anti thought experiment comes from the Pythagorean Archytas who proposed the following dilemma;

---

41 Norton uses the term anti-thought experiment to describe the potential to have two thought experiments that appear to yield contradictory results.
42 Norton (2004) p.45
43 Norton (2004) p.46
44 Norton (2004) p.46

if the universe is not infinite, then if you were to reach the end of the universe, it must either be that you can stick your hand past the edge, or that you cannot. If you could stretch your arm past the edge of the universe it would be outside the universe, which is absurd, and it is equally absurd to imagine that you would be unable to stretch out your hand. Thus the universe cannot be finite.[45] When pairs of thought experiments can be found that support opposing conclusions, Norton calls for an epistemology that will give reason to prefer one or both of the pair, and *to explain why one or both fail*. Moreover, Norton claims that appealing to outside knowledge to show which thought experiment failed is not good enough. His desideratum is to understand what went wrong "..in the failed thought experiment itself."[46] Such an epistemology would allow us to separate good from bad thought experiments and have confidence in the good ones even when their results run counter to information from another source. Norton proposes understanding thought experiments as arguments as a means of creating such an epistemology.

If thought experiments are arguments, then it is clear that they cannot be sources of knowledge equal to real experiments. Moreover, the process of running a thought experiment becomes largely irrelevant, because insofar as we have satisfactory understanding of how arguments function, thought experiments can be subsumed within the larger category. My main argument against this position is to advocate a mental modeling theory of thought experiments, which is non-propositional[47] and involves

---

45Norton (2004) p.46
46Norton (2004) p.45
47It might be argued that arguments need not be propositional, however denying that arguments need sentences which can be assigned truth values risks stripping the term argument of any meaning beyond

quasi-sensory manipulation and discovery within the process of running a thought experiment that is foreign to argumentation. If mental modeling is correct, then thought experiments are not arguments. I will begin by giving an argument against Norton that does not rely on mental modeling, and in the next chapter I will argue for a non-propositional form of mental modeling.

The idea that thought experiments are arguments has a number of advantages. Arguments are relatively well understood and widely used. By accepting that thought experiments are just arguments, the epistemic questions are answered—they have the same epistemic status as arguments. This also eliminates the questionable elements of thought experiments.[48] There is no need to look for a priori truth, or to peer into Platonic realms to understand how thought experiments can create knowledge. I will argue that mental modeling sidesteps this metaphysical complication, but it admittedly does so using theories of mental cognition that are far from uncontested. The greatest strength as I see it of Norton's argument is also its primary flaw. He claims that all thought experiments can be reconstructed as arguments. I will not contest this point, as the ability to reconstruct thought experiments as arguments alone does not make thought experiments arguments.[49] And to lay the field of arguments open to unmediated experience would be to make the concept of argument so broad as to be meaningless.[50] That there are thought experiment, anti- thought experiment pairs shows that thought

---

being anything that convinces.

48Brown (2004) p.35

49In the chapter "Why Thought Experiments Transcend Empiricism" Brown does deny that all thought experiments can be reconstructed to fit the pattern of empirically supported premises leading to a conclusion. He does so using visual and geometric examples, and by drawing on mathematics.

50Cooper p. 332

experiments can go wrong; however, that an epistemic device can be misleading is not a reason to discard it. Experiences, sensory perception, and empirical findings are all subject to error. In the following chapter on mental modeling I present two views of how thought experiments could work, using mental modeling. Both rely on a form of cognitive perception that is non-propositional and so contrary to a view of thought experiments as arguments.

Conclusion

In this chapter we have seen some of the characteristics of thought experiments in science, notably that thought experiments are like real experiments, can produce knowledge without data, and that there are different kinds of thought experiments. Brown's taxonomy included destructive thought experiments, which counter a positive claim, mediative, conjectural and direct thought experiments each of which advances a claim of their own, and Platonic thought experiments which counter an existing theory and propose an alternative in one fell swoop. The question of how thought experiment and anti-thought experiment pairs can be understood is an interesting and relevant one, and the attempt to classify thought experiments as arguments is one attempt to resolve the problem of error in thought experiments. In the following chapter I will look at two mental modeling accounts of thought experiments that explain how thought experiments can be like real experiments, and give us reason to think that thought experiments neither are arguments nor rely on a priori knowledge.

Chapter 3 Mental Modeling

Introduction

In this chapter I will show that mental modeling is the approach which best explains the process of thought experiments. The process is distinguished from the question of how thought experiments work epistemically, though the epistemic question will be clarified by an account of how thought experiments are done. By “process” I mean to identify the mechanics of how thought experiments work—of how it is that we perform them. The process of running a thought experiment will help to illuminate the different epistemic functions that thought experiments have in the production of knowledge.

I present two views of mental modeling and argue that a mental modeling approach is the best account of the process of running thought experiments. I will draw on two mental modeling theories of thought experimentation by Nancy Nersessian and by Nenad Miščević. There are significant differences in their approaches, but the relevant similarity is that both treat thought experiments as means of producing knowledge, and situate this capability in the human cognitive faculty for creating and using mental models. I will lay out Nersessian's and Miščević's views, then argue that significant differences underline the wide variety of possibilities of non-propositional reasoning suggested by both authors. Finally, I will look at a third version of mental modeling, by Tamar Gendler. Gendler's account applies a theory of mental modeling to ethics by combining the explanatory and motivational power of thought experiments. The strength of this account is that it extends mental modeling beyond scientific thought experiments which are mechanical, visual and spatial. By focusing on the schema, a mental construct that is more like a flowchart and less like a manipulable object, ethical concepts and concerns are included in the

explanatory power of mental modeling.

I will not adopt Nersessian's, Miščević's or Gendler's account complete in all specifics, as I believe that mental modeling is the best explanation but that we do not have enough information to make definitive claims about the mechanics of mental modeling at this point in time. Nersessian's account has a flexibility and a wider scope for what mental modeling is, which resonates with my subjective experience of the process of running a thought experiment. Miščević's use of chunked information[51] gives a more compelling vision of how we use existing knowledge to create mental models. In looking at both views I will be looking for a mental modeling account that describes the process of running a thought experiment. The argument I am making is that mental modeling is the best account of how thought experiments work, and of how we do thought experiments.

Not an Argument

Mental modeling theories argue that people by their very nature have the capacity for modeling. Moreover, the use of these models is non-propositional, which Nersessian uses as a response to Norton's claim that thought experiments are picturesque arguments.

> Briefly, my hypothesis is that what distinguishes thought experiments from logical arguments and other forms of propositional reasoning is that reasoning by means of a thought experiment involves constructing and making inferences from a mental simulation. This is what makes a thought experiment both "thought" and "experimental". The original thought experiment is the construction of a dynamical model in the mind by the scientist who imagines a sequence of events and processes and infers outcomes. ... While thought experimenting is a truly creative part of scientific practice, the basic ability to construct and execute a thought experiment is not exceptional. The practice is highly refined

51 Chunked material refers to Miščević's idea that knowledge and experience are stored in such a way that entire areas of knowledge can be pulled into a model, organized and self contained but usable to create, evaluate and manipulate mental models.

> extension of a common form of reasoning. It is rooted in our abilities to anticipate, imagine, visualize, and re-experience from memory. That is, it belongs to a species of thinking by means of which we grasp alternatives, make predictions, and draw conclusions about potential real-world situations we are not participating in at that time.[52]

Nersessian's theory includes analogical and visual models as well as thought experiments, and she aims to show that "these heuristics are not ancillary, dispensable aids to thinking-while the 'real' reasoning takes place by deductive or inductive arguments-but are reasoning methods essential to the practice of science."[53] Contra Norton, the act of creating or using mental models is not simply a shortcut for reasoning about geometries. It is an important part of human cognition and equally capable of producing knowledge; indeed, as she says, it is an important part of the scientific enterprise. Modeling is a different way of reasoning, not subordinate to logical or inductive reasoning.

Miščević's version of mental modeling works in two ways—via the use of a lifetime's expertise in the organization of experience, and through visual modeling that allows mental manipulations. One reasons non-propositionally; having constructed the model one then reasons *in the model*. The model is a representation that is available to manipulation in the mind the way you can manipulate a shape in Computer-aided Design (CAD) programs; having been created, it exists as an 'object' that is spatial, visual, and concrete which allows for non-propositional reasoning of exceptional power. "Although the basic mechanism enabling the construction and manipulation of the representations might be computational, the most important feature of representations and operations is precisely their concrete and quasi-spatial character. This has earned them the name of

---

52Nersessian p.292
53Nersessian p.292

'mental models'."[54] Miščević identifies three qualities to be found in mental models. First, problems involving spatial relations are easier and faster to solve using mental models than propositionally, a claim which we will see is echoed in Nersessian's claims. The focus for Miščević is on spatial problem solving, but he does leave room for temporal and other models to possess heuristic power. Second, using mental models is a concrete, specific experience with a spatial character. Third, the model is visual. Miščević is very firm about the visual nature of mental models, writing that, "The solution of the problem is 'imaged' (or 'seen'), and this is just due to the peculiarity of the medium."[55] For him mental models must be visual: "One can distinguish simple static 'frames' representing relations between a set of objects, temporal models consisting of sequences of such frames, kinematic models which is the temporal model with continuous time, and finally dynamic models which model causal relations.... The experimenter's 'view' of the model is his 'image' of it, a perspectival representation of the model."[56] This visual and spatial perception seems particularly apt for mathematics, geometry and physics, and gives a hint as to how thought experiments can solve problems in these fields.

The Mind's Eye

That mental modeling is visual is a point made by both Nersessian and Miščević; however, the sense in which it is visual for Nersessian is only that she locates the process in the visual cortex, which is a matter for empirical refutation or confirmation and not of special concern in my account, and that the form of reasoning found in mental models is definitionally distinct. What is required is that there be a model of a situation constructed

---

54Miščević p. 220
55Miščević p. 221
56Miščević p. 220

in the mind which is open to changes and manipulations, not that this be understood or represented as pictorial. Nersessian writes that “I want to stress that most researchers would concur in the view that mental modeling, even if it does make use of the mechanisms of the visual cortex, is not like constructing a picture in the mind. That great thought experimenters, such as Bohr, have claimed not to be able to visualize well does not undermine my claim that thought experimenting is mental modeling. Mental modeling does not require introspective access to an image in the "mind's eye". It only requires the ability to reason by means of an analog model.”[57][58] This is a distinct advantage for Nersessian's view because it maintains the important features of a model--that you can work with it and that elements in the model have stability that allow for such manipulation--without committing to a visual-sensory analogue which may prove restrictive. Just as a blind person might use a physical model, and be able to manipulate the apparatus without having visual access, an explanation of mental modeling which denies analogues that match tactile, auditory, temporal, and possibly olfactory or taste sensory perception limits itself unnecessarily. If we take thought experiments to be like real experiments, and mental models to be the experimental apparatus, it is forming the model and using it that is of interest. Moreover, making visualization a necessary feature of mental modeling is problematic if we want to include thought experiments in ethics. While you may easily picture a runaway train about to run over five innocent bystanders, it is harder it envision justice or the good. By placing such emphasis on the physical and mechanical aspects of mental modeling, we risk losing the most important elements of

---

57Nersessian p.294

58I will look more closely at the use of analogical models in chapter five, where I apply Mary Hesse's theory of analogical arguments to issues of possibility in using science fiction thought experiments.

philosophical thought experiments. The colour of the runaway train is less important than the idea of unintended consequences, but we can picture the former with ease and the latter only obliquely.

Visualization is central to Miščević's theory of mental models. He describes the ability to manipulate models as a visual talent, without which mental models are not possible:

> The manipulability and the mobilizing force should account for the heuristic value of thought experiments. The theory of mental models makes correct prediction that a visualizable thought experiment will have more heuristic value than a corresponding propositional alternative, and explains in principle why. It is plausible that manipulating quasi-spatial models can be equivalent to *actually seeing* the movement of pieces manipulated. The manipulation is quasi-experiential in that it yields non-inferential beliefs akin to perceptual ones. This might explain the *felt* difference between thought experiment and a piece of regimented propositional reasoning.[59]

In this account the visual modality is essential for the ability to manipulate mental models. In turn the ability to mentally manipulate explains the ease with which problems can be solved using mental models. Thus the visual nature of mental models is central for Miščević's theory as a whole. This opens the theory to critical claims or evidence that those who can conduct thought experiments might not visualize them. If it is true that mental models can be used in other sensory modalities, then Miščević's theory is undermined.

In the relative importance of long- and short- term memory in mental modeling we see another difference between the accounts of Miščević and Nersessian. One major difference in their approaches is that Nersessian explicitly locates thought experiments as constructs of working memory, whereas I believe that Miščević's theory places greater

---

59Miščević p. 224

emphasis on access to structures that persist in long-term memory. Nersessian makes the following distinction: “The most significant distinction for our purposes is between those investigations that treat mental models as structures stored in long term memory and then called upon in reasoning and those that treat them as temporary structures constructed in working memory for a specific reasoning task.”[60] Nersessian explicitly locates her theory in working memory; however, there is significant overlap with Miščević's theory as both take experience and prior learning as determinants of how successful mental models will be; it is the existing knowledge of systems that allows accurate models to be constructed. The crucial difference is that Nersessian firmly situates her mental models in working memory, whereas Miščević does not so specify, and places more weight on the experiential background. He writes that “Although mental models are based in part on static prior knowledge, they are themselves transient, dynamic representations of particular unique situations... mental models are the major source of inactive change in long-term knowledge structures. “[61] In considering models transient, the working memory is implicitly involved, but Miščević's focus is on using long-term knowledge to build models, and the effects models have on long-term memory. I believe that this makes the part of thought experiments that occurs in the working memory simply procedural—a necessary means of mental processing, and not the home of mental models. We can see Miščević's emphasis on the importance of long-term memory in building models in his use of everyday knowledge which is, of course, part of our long term memory.[62] Nersessian's theory takes existing representations and through narrative

---

60Nersessian p. 293
61Miščević p. 221
62It has been pointed out to me that what counts as everyday knowledge may be up for grabs in the context

constructs a model in the working memory: “In constructing and updating a model, the reader calls upon a combination of pre-existing conceptual and real-world knowledge and employs the tacit and recursive inferencing mechanisms of her cognitive apparatus to integrate this with the information contained in the narrative”.[63] At the end of this chapter I will dwell in more depth on the role of narrative on Nersessian's theory.

Back to Galileo

In Chapter One I gave examples of some of the impressive things that thought experiments are able to do. In using a mental modeling account to show how thought experiments map onto the world and enable rapid and accurate judgements, I believe we have the beginnings of an answer to the epistemic question of how thought experiments produce knowledge without the addition of new data—mental modeling enables cognitive capacities and the expression of background knowledge. For an example of this, I turn to Miščević's account of the now familiar example of Galileo's falling bodies. Miščević offers an account of how mental modeling can explain the working of this thought experiment:

> In terms of mental models this is how he does it: He first builds two smaller models for L and H; first an 1-model, with L falling slowly; and an h-model, with H falling rapidly. Now, he is asked to consider how H and L would fall together. This demands an *integration of h-model and l-model.* When he tries to perform the integration he discovers that an integrated lh-model, which would be at the same time model for P, is impossible. Instead of manipulating formulae (the way a logician would do) Simplicio manipulates more concrete representations

---

of science fiction (Thanks to my supervisor, Letitia Meynell, for raising this issue). I would respond that in science fiction those elements that are meant to contravene or be outside of our experience are spelled out, and in all other respects the reader is expected to use common sense to fill in what is not specified. Thus in a story about posthuman cloud computed entities who temporarily instantiate themselves in created bodies, we still expect that if their created forms want to go in a door they open it first.

63Nersessian p.294

> enabling him to use his everyday skills and will certainly reach with ease the required conclusion (the interim conclusion that it is incoherent to maintain that speed is proportional to weight). The purely propositional way might have been, on the contrary, closed to Simplicio.[64]

I believe the most important element of this expression of Galileo's thought experiment in terms of mental models is that the manipulation of the models is representative and allows for interpretation through a comparison that is more like looking at diagrams than like using mathematical formulas. The creation of models allows for the use of cognitive skills that are found not in propositional reasoning but in spatial reasoning. The use of background knowledge in creating mental models enables the positive claim that all bodies fall at the same rate. The reason is that "...truths about what is irrelevant are given as implicit constraints on our model building. They determine the architecture of the model, while remaining rather inconspicuous in the background."[65] That the colour, taste, appearance, or chemical composition of an objects would affect the speed of an object's fall does not accord with what we know about the world, and so our representation of the thought experiment excludes these options:

> All the irrelevant possibilities (irrelevant from the viewpoint of the enquirer) are excluded thanks to the filtering power of default hierarchies. ... together with this stratified knowledge, the use of models also mobilizes knowledge embodied in enquirers cognitive skills, most notably spatial ones, all contributing to the specificity of the model, narrowing the range of the alternative possible world to make it as close as possible to (what the experimenter takes to be) the actual world.[66]

This stratified knowledge is important for Miščević's concept of chunked material.

64Miščević p. 222
65Miščević p. 225
66Miščević p. 225

Miščević theorizes that by using our general knowledge of the world we create 'default hierarchies', which are the sum of our experience, and naturally fall into play unless otherwise specified: “The building of default hierarchies is usually presented in terms of inductive learning, in the very broad sense of 'induction'. Alternatives are tried and rejected, rules (conditionals) are recombined and transformed, their 'credibility' being subject to empirical checking. In this way, the finished default hierarchy embodies results of a long learning process.”

Chunked material refers to schemas and scripts that make knowledge and past experience available. The method that Miščević proposes for this process is that of default hierarchies. Our background knowledge is “...chunked, with undetachable 'parts', stratified in default hierarchy and medium-specific.”[67] Miščević applies this idea of chunked material to Galileo's falling bodies, writing that

> Your knowledge is organized (and stored) in a default way and it might be present implicitly in various scenarios concerning situations of falling. It is difficult to handle outside a concrete context (try to think of all the things that you in fact know to be irrelevant for the speed of a falling body). If one had to write down a list, and to explain why the particular characteristics are on the list, it would be a never ending job.[68]

Identifying salient features is normally fluid and effortless given concrete situations, and modeling enables this function; default hierarchies are set in motion when mental models are created and manipulated, and so control and constrain the model as simply as they do our understandings of real situations. In explaining Galileo's falling bodies Miščević gives an example of the two elements of his theory that I consider provide the

---

67Miščević p. 225
68Miščević p. 225

greatest explanatory power; that the use of mental models allows for visual and geometrical reasoning, and that default hierarchies allow for the use of chunked material. Thus the production of knowledge is explained by a powerful and distinct form of cognition and the harnessing of experience through the use of default hierarchies and chunked material. The creation and use of default hierarchies is also implicated in Miščević's explanation of the accuracy and rapidity of reasoning using mental models in the following section.

I Can't Believe it Worked So Well

Both Nersessian and Miščević address the rapidity of solutions using thought experiments. The question of how we can reason so quickly and accurately using thought experiments is presented both to support the usefulness of thought experiments and to show that mental modeling is the best account.

Geometry is taken up by Miščević who acknowledges that drawings are widely used in teaching and practising geometry because it is simply easier to understand some problems of geometry visually. He attributes this ease to spatial reasoning, saying that "Everybody knows in practice which properties a circular chain retains while sliding. But unless one has been doing some topology, one has no way to express this knowledge systematically in a simple propositional way... However, when you build your model or Stevin builds his, you use your spatial skills—the same ones you use when orienting yourself in space (or playing chess or tetris)."[69][70] I would say that this is done in the same

---

69Miščević p. 223

70Simon Stevin created a thought experiment in which a chain is laid over a double plane that is entirely frictionless. The idea is that unless we start with static equilibrium we would have created a perpetual motion machine.

way that seeing an illustration showing examples of Equilateral, Isosceles and Scalene triangles is easier to grasp than being told that the number of equal angles defines these categories.

Nersessian uses this speed of processing as support for her theory, arguing that "Some of the experimental evidence for the hypothesis of mental modeling during narrative comprehension comes from chronometric studies which claim to show that model-based reasoning is faster than reasoning with propositions. A situation that is represented by a mental model should allow the reasoner to generate conclusions without having to carry out the extensive operations needed to process the same amount of background information to make inferences from an argument in propositional form."[71] I believe that this is analogous to the ease of catching a thrown ball, and the relative difficulty of calculating the intersecting vectors of your hand and the ball. Rather than arguing for the efficacy of thought experiments, Nersessian is arguing that because there is evidence that we do reach correct solutions faster for some problems by using mental models, an explanation is called for, and the best supported explanation is that the ability to use and adapt background assumptions, to hold a model of the relevant circumstances, and to manipulate this model with its associated constraints, best explains the speed of reasoning. She writes that,

> The situational constraints are built into the model, making many consequences implicit in it that would require considerable inferential work in propositional form. For example, moving an object changes, immediately, its spatial relationships to all the other objects. The reasoner would grasp this simply by means of the changes in the model and not need to make additional inferences. Further, reasoning through a model should restrict the scope of the conclusions drawn. For

71Nersessian p.295

> example, moving an object in specified manner both limits and makes immediately evident the consequences of that move to those directly relevant to the situation depicted by the narrative. Thus other support comes from demonstrations that inferences subjects make are much more difficult or not made at all when they are required to reason with the situation reformulated propositionally[72]

It is relatively fast and easy to use a physical model to figure out what will happen if you press the lever or drop the weight because all of the elements that would need to be tracked and accounted for are already part of the model, and thus it is only the results that require attention. If mental models can be used in a relevantly similar fashion, the speed and accuracy of results could be explained in like manner, and arguing from the conclusion does not of course show that this is true, but just that it fits the facts that are not otherwise explained.

Brown offers an alternative account of the effectiveness of thought experiments that relies on thought experiments' ability to tap into the Platonic realm of truths. Miščević structured his article "Mental Models and Thought Experiments" as a response to Brown's Platonist account of thought experiments. I present Miščević's account as a means of explaining the distinguishing features of thought experiments, which I believe are better accounted for by mental models than by appealing to *a priori* access to truth. Miščević makes the claim that mental modeling makes sense of thought experiments as non-apriorists have struggled to do, thus making *a priori* access to truth unnecessary for an account of mental models. The four points that Miščević sets non-apriorists the task of responding to are:

72Nersessian p.295

> 1. How can one learn new things without new observational data?
>
> 2. Why are thought experiments superior to deduction in terms of heuristic power, obviousness and ease?
>
> 3. Where does the 'experiential' element in thought experiments come from? Are there any new experiences or quasi-experiences present in thought experiment, and of what nature are they?
>
> 4. If the reasoning in thought experiment is broadly inductive, how can it eliminate alternatives and reach its conclusion so quickly and effortlessly, and assert it with such force?

Miščević argues that mental modeling answers these four questions: "It explains the nature of data, points to the skills employed in manipulating the items in the model, and accounts for heuristic fruitfulness and relative ease of the work in the laboratory of the mind"[73]

Another feature that Brown identified as Platonic is that thought experiments can supersede the need to do real experiments even when it is possible. As I described in relation to the Salviati ship example, there may be reasons to prefer the thought experiments. Nersessian takes a similar view saying that "A thought experiment is usually so compelling that even in those cases where it is possible to carry it out, the reader feels no need to do so."[74] She goes on to justify this statement by arguing that the model is a sufficiently accurate representation of the world that we can make claims without the need to perform the experiment on a real model. This linkage of mental models to the external world is an important justificatory element in a mental modeling account of thought experiments; to be able to claim that thought experiments really show us something, mental models lack recourse to *a priori* knowledge and so must give an

---

73Miščević p. 226
74Nersessian p.296

account of how they not only model the situation, but do so in a way that gives us knowledge of the actual world and not just our mental processes. But conversely they have as a strength that they need not assume a Platonic Realm.

Narrative in Mental Models

In this section I look at Nersessian's account of narrative, and attempt to remove a few future pitfalls[75] for my argument that literary thought experiments are effective, and particularly that science fiction thought experiments are apt in bioethics. In particular I wish to ally mental modeling and thought experiments in literature. In the following chapter I will argue that fictional narratives can be thought experiments. Thus it is interesting to note that Nersessian proposes that the narrative form of thought experiments guides our reasoning. This is especially interesting because of her vehemence that we understand thought experiments in a non-propositional manner; Nersessian argues that the presentation of a narrative encourages us to create mental models or schemas rather than a set of propositions. Nersessian writes that "Reading, comprehending, and thinking about stories would seem to epitomize thinking with language. Yet, there is a significant body of cognitive research that supports the hypothesis that the inferences subjects make are derived from constructing and manipulating a mental model of the situation depicted by the narrative, rather than by applying rules of inference to a system of propositions representing the content of the text."[76] Nersessian is using a concept of mental models that is relatively broad, and

---

75By establishing that narrative is an important element of thought experiment, I propose to ease the move to literary thought experiments. If thought experiments are narrative in nature, and fictional (counterfactual) it is not as great a leap to say that literature can be thought experiments, as it is also narrative and fictional.

76Nersessian p.293

includes not only spatial pictures like geometrical figures that we can mentally rotate, but also the schemas we create for everyday activities. Notably, Miščević also writes that narratives are envisioned using mental models. He explains, "When a reader encounters a description of a situation, she builds a model, a quasi-spatial 'picture' of it. As new details are supplied by the story-teller, the model becomes updated. The background conditions are dictated by the reader's general knowledge about the world."[77] Thus even stories which are not thought experiments are the source of mental models in the theories of both Nersessian and Miščević. I will focus on Nersessian's account of narratives in thought experiment.

The narrative form of thought experiments makes them easier to share and understand. Nersessian identifies a pedagogical role for narrative in thought experiments, namely that "Once the initial experimenter understands the implications of a thought experiment, she can guide others in the community to see them as well by crafting a description of the experiment into a narrative."[78] We hold onto stories, remember them, use them to understand, and it is this structural organizational and memorable nature of stories that I believe Nersessian makes use of here. Nersessian has written extensively about the role of narrative in transmitting the ideas that are found in thought experiments,[79] and the idea that by making ideas easy to remember and to share, the narrative form of thought experiments increases the ability to transmit ideas. Presumably, this acts in concert with the increased ease of problem solving that accompanies creating models, thus adapting

---

77Miščević p. 220
78Nersessian p.299
79"As I have discussed elsewhere (1991a, 1991b, 1992), the narrative form of presentation plays a central role in communicating a thought experiment within a community of scientists" (Nersessian p. 291)

thought experiments well to sharing ideas. Narrative transmission is relevant to bioethics as policy change is one of the possible purposes of bioethical debate. If it can be argued that thought experiments are useful means of transmitting information, then this gives an additional reason to use thought experiments in bioethics. In the next chapter I will give reasons to accept thought experiments in bioethics, and one reason is to provide a tool for reasoning about future developments. Because it is harder to make policy and regulatory changes once new products are on the market and practices are in place, the use of thought experiments as means to understand and communicate new developments is very useful.

One element of Nersessian's account seems clearly opposed to the use of literary narratives as thought experiments. That is the requirement that thought experiments refer to abstractions in order to make the thought experiments prototypical.

> The thought-experimental narrative depicts abstractions. For example, certain features of objects that would be present in a real-world experiment are not included, such as the color of rocks and the physical characteristics of observers. That is, there has been a prior selection of the pertinent dimensions on which to focus that evidently derives from experience in the world. From experience we believe, e.g., that the color of a rock does not affect its rate of fall. Such information is customarily excluded from real-world experimental narratives as well. This facilitates the reader's recognition of the situation as prototypical, i.e. as representing a class of experimental situations.[80]

Fiction is unlike thought experiments in that it is often more elaborate—more details are filled in.[81] Does this account then make fictional narratives impossible to generalize from, and so incapable of acting as thought experiments? I would argue that this is not the case

---

80Nersessian p.296

81 Thanks to my third reader, Michael Hymers, for pointing out that some fiction, such as "Two Meditations" by John Barth, which is less than a page long will have fewer details than thought experiments in philosophy. This will likely also apply to some poetry and short stories.

because we are very good at making stories prototypical. When we hear Aesop's fable of the tortoise and the hare, the message we pick out is not that *that particular hare* was overconfident. We instead take the general message that persistence is a good thing. I believe that fictional narratives need to be generalizable in a specific way; we cannot believe that the issue at hand is only possible, or that the solution works only in the particular situation presented. The reader must be able to see more general application for the conclusion that is reached. Nersessian writes that embroideries which are strictly speaking extraneous may serve as emphasis:

> It is true that, as John Norton (1991) has pointed out, extremely colorful narratives may include highly specific details. Rather than being "irrelevant", as he maintains, though, these details usually serve to reinforce crucial aspects of the experiment. For example, in one version of the chest, or "elevator", experiment, Einstein depicts the physicist as being drugged and then waking up in a box. This colorful detail serves to reinforce the point that the observer could not have known before entering the chest if he were falling in outer space or sitting in a gravitational field. It also reinforces the condition that the observer cannot know whether or not there are gravitational sources around.[82]

It seems to me that not only can the occasional detail add emphasis, but so can a fictional narrative contribute to the memorability and the context of a thought experiment without ruining the generalizability of the thought-experimental results. It is also worth questioning why we believe that thought experiments must be prototypical. The reason Nersessian gives is that it they can thus represent a *class.* The virtue of having the thought experiment stand in for a class of experimental situations is that it is then justified to make a universal principle for that class. But given the uncertainty over whether there are moral truths or laws of nature, this may be an overly ambitious project,

---

82Nersessian p.296

and even an unnecessarily ambitious one. Certainly a thought experiment that has no wider applicability is either useless or not a thought experiment, but asking for universal principles may be asking too much, when all we really need is a field of application for the thought experiment in question. In the final chapter I will adapt Mary Hesse's account of analogical models to describe relevant possibility. This account may also apply to the generalizability of literary thought experiments by showing that only relevant aspects of literary fiction must have cognitive upshots.

Thought experiments are like other narratives in that they rely on our background assumptions about the world to fill in all of the details that are not included. In a story it is generally assumed that unusual elements will be described, and that what is not mentioned will be in essence what the reader expects of the world. An author might give a detailed account of how they propose gravity will work on their imagined space station, but when they mention a character has a drink of water we picture this in the familiar ways that we would get a drink of water. Nersessian writes that "The reader is invited to follow through a sequence of events or processes as one would in the real world. That is, even if the situation may seem bizarre or fantastic, such as being in a chest in outer space, there is nothing bizarre in the unfolding. Objects behave as they would in the real world in the presence or absence of gravity. The assumption is that if the experiment could be performed, the chain of events would unfold according to the way things usually take place in the world."[83] This characteristic of narrative forms is allied by Nersessian with another characteristic of stories, which "[unfold] in time and [follow] a specific causal

---

83Nersessian p.295

sequence."[84] In this section I have argued from the claim that thought experiments are narratives, and that narratives allow us to create mental models. I have used these claims to pave the way for the idea of literary thought experiments by arguing that mental modeling explains the appropriate use of background knowledge in reading literary narratives, and that extraneous details do not stop us from generalizing from literary narratives. These features give us reason to think that thought-experiments and narratives are similar, and indeed Nersessian has claimed that they function in much the same way. If it is true that thought experiments are in this robust sense (and not merely in form) narratives, then what I need to argue is not that some narratives can work as thought experiments, for thought experiments are narratives, but rather that the narratives I am interested in, namely science fiction that is about issues in bioethics, are thought experiments.

Mental Modeling to Explain and Motivate

In "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium", Tamar Gendler introduced a mental modeling theory for ethical thought experiments. In this section I will describe this view, suggest that it could do more to account for moral intuitions and finally argue that it sufficiently accounts for the explanatory and motivational powers of ethical thought experiments. Gendler writes that "Viewed in this light, moral and political philosophy have a secondary task that runs alongside the task of ascertaining what morality demands, namely, that of providing the reader with resources that enable him or her to make the perspective shift that the moral stance requires at the moment of moral decision-making. ... images that will bring the readers to reframe their

84Nersessian p.295

experience of some morally valenced situation, in such a way that their apprehension of the morally relevant features of it are re-experienced in light of the scenario presented. It is this role, I want to suggest, that is played by some of the most famous thought experiments in moral and political theorizing."[85]

Gendler's account is both explanatory and motivational, as the thought experiment is used to parse the relevant moral aspects of the situation, and then this shift in perspective acts as motivation to make the ethical decision thus indicated. Gendler locates the power of thought experiments in the abilities of a theorized experiential system, which "operates in a holistic, associationist manner, is intimately associated with the experience of affect, represents events in the form of concrete exemplars and schemas inductively derived from emotionally significant past experiences, and is able to generalize and construct relatively complex models for organizing experience and directing behavior by the use of prototypes, metaphors, scripts, and narratives." [86] Gendler's theory of thought experiments is a mental modeling theory, but it is distinct in placing the emphasis on schemas rather than visual representations.

Schemas figure prominently in some strains of cognitive behavioural therapy, and psychological understandings of cognition. Schemas are ways that people organize their experiences. They enable people's ability to easily distinguish what kind of situation they are in, and what responses are open to them. Thus you might have a schema about grocery shopping which is activated when you walk into Sobeys, and which includes a series of movements through the store and interactions with other shoppers and store staff.

85Gendler p. 83
86Gendler p. 75

Schemas may seem very far removed from mental modeling theories which are visual or spatial, and emphasize manipulation of modeling elements, but you could equally well have a schema for how a lever works which is very much like a mental model in that it suggests a causal chain resulting from your application of force in a number of ways. What is very different is that schemas are not necessarily visual, and are more like scripts and less like experiments. Gendler connects the situated or enacted characteristic of thought experiment with schemas discrete from purely abstract propositions.

> For our purposes, the moral is simply this. Decades of research in cognitive psychology have demonstrated that when content is presented in a suitably concrete or abstract way, this may result in the activation or fortification of a representational schema that was otherwise inactive or subordinate; the result of this may be to evoke responses that run counter to those evoked by alternative presentations of relevantly similar content. So far from being an anomalous or idiosyncratic feature of arcane or unusual cases, the discrepancy described in our opening story is—in fact—a central feature of our mental lives.[87]

This differential activation of schemas is intended to explain situations in which our theoretic moral commitments run counter to our moral intuitions. A strength of Gendler's account is that it makes sense of conflicting responses that we experience, and of the question of what it means for moral intuitions to be in conflict with moral commitments. What I do not necessarily agree with is Gendler's view that it is moral intuition that is wrong.

Gendler presents the Wason selection test in which subjects are told that their goal is to falsify a statement of the form 'if a given card has the letter A on one side, it will have the number 3 on the other'. Subjects are then shown four cards, each of which is printed with one of the following; the A, B, 3 or 7. To show that this statement is false they must

---

87Gendler p. 75

select the A and the 7 card.[88] Subjects often chose A and 3, and a number of studies have been made to attempt to discern the source of this cognitive error. “Like optical illusions, these cognitive illusions seem to be artifacts of deep features of our cognitive architecture: ... Just as we cannot simply talk ourselves out of seeing Müller-Lyer lines as different in length, we cannot simply talk ourselves out of feeling drawn toward turning over the 3.”[89] Similar experiments using syllogisms find that errors are much more likely when there is an invalid inference that has a clearly true conclusion, which gives reason to think that independent judgements affirming the truth of the conclusion out-compete attention to the structure of the argument. This cognitive error is likened to the operation of moral intuitions by Gendler, who writes that “...if something akin to dual processing lies at the root of most human reasoning, then a philosophical theory may be correct even if we consistently and resiliently react to specific cases in ways that run counter to the theory’s predictions.”[90]

Even if Gendler is entirely correct about dual processing, which is a psychological theory used to account for how a phenomenon can occur in two different ways, it is unclear what grounds that gives for saying that moral intuitions are incorrect and moral reasoning correct. All that the existence of two different layers of thinking shows us is that we may reach different results by looking at an issue in a concrete or abstract manner. Gendler admits that moral intuitions may sometimes be relevant, however this is contrasted with cases in which Gendler believes they are flawed due to the inclusion of morally irrelevant features;

88Gendler p. 76
89Gendler p. 76
90Gendler p. 76

> All of this is fully compatible with there being a genuine deep moral difference between the two acts—deep enough to render the one morally mandatory and the other morally prohibited. Nothing that I have said here or elsewhere should be taken to deny the possibility that—as Mill writes at the beginning of Utilitarianism—"whatever steadiness and consistency our moral beliefs have attained has been mainly due to the tacit influence of a standard not yet recognized" (Mill [1861] 2001, 3). That said, it is worth taking seriously other work that suggests that intuitions about such cases may vary along dimensions that are (presumably) completely morally irrelevant.[91]

When moral intuitions are linked to the cognitive mistakes that people make in reasoning problems, it does seems clear that 'non-moral' elements, similar to framing problems,[92] cause error. But it is worth taking some care in what we consider to be non-moral. Gendler talks about how 'not relevant' facts affect moral intuitions in thought experiments, but she does not show that this is restricted to thought experiments, or even that they are completely morally irrelevant. One of her examples has participants over-correct for their own racism by disproportionately agreeing to sacrifice educated Caucasian people to save people from a poor black neighbourhood. If these are not empty words, spoken with the goal of not appearing prejudiced, there may in fact be moral reasons to think that there is a greater duty to save the oppressed members of society, and in particular the historical context of slavery may give those who benefit from the continued structural injustice of this historic injustice a special moral obligation. You need not agree with this claim, only admit the possibility that the 'non-moral' features of this experiment may in fact be morally weighted. But even when the influences are non-moral, they are shown

---

91Gendler p. 77

92Framing problems are of course not unique to moral intuitions. In the article "Mental Models and Thought Experiments" Miščević wrote the following about scientific thought experiments; "It is here that we face a problem—a variant of the famous 'frame problem': how does the enquirer select what is pertinent? How are (what are considered to be) the irrelevant factors kept at distance, so that they don't overburden the argument? How is the information about all the relevant factors assembled?" (Miščević p.218)

only in examples that are unclear or involve dilemmas. And if what we have learned is that in borderline cases our intuitions can be influenced by irrelevant factors, this may be less than conclusive as to the value of moral intuitions. In the following chapter on thought experiments in ethics I will consider the question of moral intuitions in more depth.

While I do not agree with the marginal role that moral intuitions play in Gendler's account, her use of schemas to evoke responses that lie dormant in abstract or propositional reasoning is of great interest and, I believe, convincing. Gendler writes, "that by presenting content in a suitably concrete or abstract way, thought experiments recruit representational schemas that were otherwise inactive, thereby evoking responses that may run counter to those evoked by alternative presentations of relevantly similar content."[93] One concern that is suggested by this view is in identifying what it is about thought experiments that makes them concrete, or perhaps what kind of concreteness is found therein. Gendler uses 'concrete' as her primary description for what differentiates thought experiments. But what exactly does it mean to say that thought experiments are 'concrete'? It is hard to pinpoint what it is about a thought experiment such as that of Parfit, in which people split like amoebas, that is more concrete than a moral principle such as 'do not steal'. I have an easier time understanding what it would mean not to steal, than what it would be like to live in a world of people who split. I believe that this can be answered by Gendler, and by mental modeling theories more generally, in looking at the process of cognitively enacting thought experiments. In using a schema or model the mind must do the work of making a concrete representation, and in so constructing the

93Gendler p. 69

details must be 'filled in', moving the modeler away from simple abstractions and towards more complex imaginative representations.

Gendler's theory can be used to explain how moral thought experiments can motivate and explain. I believe that this is borne out when Gendler describes the mechanisms by which thought experiments can influence moral judgement. The following quote expresses the idea that thought experiments can create or activate schemas to understand moral content: "when thought experiments succeed as devices of persuasion, it is because the evoked response becomes dominant, so that the subject comes (either reflectively or unreflectively) to represent relevant non-thought experimental content in light of the thought experimental conclusion."[94] Here the workings of thought experiments are persuasive, and work by creating new behavioural schemas that are used to understand acts according to the thought-experimental representation. In the following quote Gendler clarifies how schemas act in an explanatory capacity: and in the following quote Gendler clarifies how a schema can act in an explanatory capacity: "It provides the subject with a powerful frame through which the target material... can be reconceptualized. It seeks to make the moral stance cognitively available at moments of moral decision-making."[95] Using a schema as a blueprint for moral situations is both a way of understanding the moral elements, and provides a set of possible responses to the situation. A schema is both a means of understanding or categorizing and a guide for action; it has both an explanatory and a motivational function. In using schemas to explain how we motivate and explain moral actions, Gendler's account gives a workable

---

94Gendler p. 69
95Gendler p. 85

mental modeling theory for ethical thought experiments. I believe that changes in her account to allow a role for moral intuitions as potentially relevant and important in thought experiments also has the potential to make a justificatory mental modeling account possible in ethics.

Conclusion

I have argued that mental modeling is the best account for the things that thought experiments can do. In addition, the narrative character of thought experiments as understood by mental modeling prima facie supports the idea that literary thought experiments are useful in bioethics. I believe that mental modeling is the best explanation, but that we do not have enough information to make definitive claims about the mechanics at this point. Mental modeling is best understood to partake of the flexibility of Nersessian's account, combined with Miščević's compelling vision of how we use existing knowledge to create mental models, and Gendler's use of schemas to understand ethical thought experiments. My view of mental models relies heavily on the use of imagination to create and manipulate models or schemas which have the characteristics of using background information, being available in different sensory modalities, using narrative and using schemas to run ethical thought experiments, as well as spatial or visual models to run scientific thought experiments.

Chapter 4 Thought Experiments in Ethics

Introduction

In this section I begin with some questions about the efficacy of thought experiments in reliably eliciting moral intuitions. First, I will look at the idea of thought experiments as intuition pumps, then I will look at what role moral intuitions play in ethical thought experiments. Next, I will consider the motivational power of thought experiments, focusing on Singer's drowning child thought experiment, and then on the explanatory power of thought experiments, in considering Thomson's violinist, and justificatory power in social contract thought experiments. Finally, I will present an account that I argue serves to position the use of explanatory and motivational thought experiments in ethics. According to the account by Hintikka, it may be seen that doing multiple thought experiments will allow us to identify moral intuitions and use them for explanatory and motivational purposes.

Broadly speaking, I argue that thought experiments in ethics function by eliciting moral intuitions. There is a distinction that I will draw between the use of thought experiments in ethics, depending on whether they are intended to have an explanatory, justificatory or motivational function. The virtue of this distinction is that the thorny question of the status of moral intuitions is less troubling in explanatory and motivational contexts.[96] I agree with Dennett that ethical thought experiments work as intuition pumps, but I will argue that in explanatory and motivational contexts this should be acceptable to us. This is not to say that none of the issues with how moral intuitions are elicited will be relevant—framing problems, lack of consensus and so on will still pose problems. I

96 My thanks to Adam Auch,for suggesting this in a conversation on June 17th.

believe it is fair to say that if the moral intuitions that are elicited using thought experiments are intended to be justificatory, then we need an account of moral intuitions in which they draw on moral truth; in order to be able to justify either an epistemic claim about what morality is, or a normative claim about what morality requires, it seems that we must first establish that there is something like moral fact and that moral intuitions give us access to these moral facts. I will for the moment allow that justificatory thought experiments are problematic, though the problem of what the underpinning is for our moral sense is problematic for the entire field of ethics and not only for thought experiments or for moral intuitions. I will then apply the taxonomy of thought experiments created by Brown, which I introduced in the first chapter, to ethical thought experiments, with an eye to how my distinction among justificatory, explanatory and motivational ethical thought experiments fits into Brown's categories. This will pick up on differences between ethical and scientific thought experiments, which will both be drawn upon in bioethics.

Problems with Intuitions

In the literature about thought experiments the term 'intuition pump' has become a shorthand criticism of thought experiments. This critique positions intuitions as useless or misleading, and stigmatizes philosophical thought experiments as doing nothing but drawing out worthless and uninformative feelings or hunches. In the introduction I introduced Goodenough as a foil, and he writes that his purpose is to show that thought experiments that draw on intuitions are problematic. He goes on to say: “But the use of thought experiments in philosophy is more problematic. They can be used to reveal contradictions implicit in assumptions and conceptual schemata, but more often they are

used as what Dennett has called 'intuition-pumps'"[97] He claims that the use of thought experiments in science and to reveal contradictions is possibly acceptable, but those thought experiments that are 'intuition pumps' are useless and dangerously misleading. Thus to make a case for the use of ethical thought experiments, I will begin by considering whether intuitions are always bad. The debate over the use of intuitions pumps is not specific to ethics, but is in relation to philosophy more generally. It may be that intuitions play such a central role in philosophical disciplines including ethics, epistemology and metaphysics that eliciting and parsing these intuitions takes on weight in philosophical thought experiments. In the first part of this chapter I will look at some pros and cons of of philosophical intuitions, and specifically ethical intuitions.

Daniel Dennett coined the term 'intuition pump', which he applied to the Chinese Room thought experiment by John Searle.[98] Dennett charged that Searle failed to make an argument, and simply appealed to our intuitions.[99] In the article "Intuition Pumps" Dennett writes the following about intuition pumps: "They're not arguments, they're stories. Instead of having a conclusion, they pump an intuition. They get you to say 'Aha! Oh, I get it!'"[100] It is important to note that Dennett himself does not condemn all intuitions pumps, and indeed affirms the use of some intuition pumps.

97Goodenough p.7

98Searle creates a scenario in which an English speaking person is placed in a room alone, with a complete set of instructions for responding to strings of Chinese characters. The intention of the thought experiment was to show that the ability of computers to respond to human languages does not imply that computers have the capacity to understand human languages any more than the person in the Chinese room understands Chinese because they can look up strings of characters (sentences) and find in their reference materials appropriate responses.

99Dennett p. 182

100Dennett p. 182

> If you look at the history of philosophy, you see that all the great and influential stuff has been technically full of holes but utterly memorable and vivid. They are what I call "intuition pumps"--lovely thought experiments. Like Plato's cave, and Descartes's evil demon, and Hobbes' vision of the state of nature and the social contract, and even Kant's idea of the categorical imperative. I don't know of any philosopher who thinks any one of those is a logically sound argument for anything. But they're wonderful imagination grabbers, jungle gyms for the imagination. They structure the way you think about a problem. These are the real legacy of the history of philosophy. A lot of philosophers have forgotten that, but I like to make intuition pumps.[101]

This description of 'jungle gyms for the imagination' shares some qualities with mental modeling theories. Indeed structuring how you think about a problem gets you a long way towards solving it.[102] I believe Dennett is suggesting that thought experiments help us build structures or scaffolds on which our imagination can play—manipulating and exploring using the (presumably) mental forms that thought experiments lead us to imagine. It would be a hard path to deny that thought experiments dredge up our moral intuitions, and indeed neither critics nor proponents of thought experiments deny this faculty. What is questionable is what kind of intuitions emerge, and what utility these intuitions have.

Are Thought Experiments Really Intuitions Pumps?

Beyond all questions of the reliability of moral intuitions, we may ask whether or not hypothetical situations elicit responses that will hold true in actual situations. Whether or not the responses of those in the actual situations and those asked to imagine that they are so situated are highly correlated is a question for empirical testing, but it seems that there will be a large area of overlap and some divergences. For an example, I believe that we

---

101 Dennett p. 182

102 Thanks to my supervisor, Letitia Meynell, for this idea

may look to the famous Milgram experiments in which experimenters told participants that they were administering electrical shocks to another volunteer, who could be seen behind a glass window. In actuality the 'volunteer' was a confederate of the experimenter, and the electrical shocks were simulated. The surprising and morally troubling finding of this study was that many participants were willing to administer what they had been told was a potentially lethal shock to a 'volunteer' on minimal urging from the experimenter. What is potentially more interesting is the large literature examining this study, because this is *not* the result that we would have expected. Most people would likely say, if questioned, that in this situation they would refuse. This points to inaccuracies in how we perceive ourselves-- disproportionately well, morally as well as along other axes, and to the gap between what we say and think we will do and the actual behaviour that people exhibit in morally relevant situations. Kwame Anthony Appiah brings up a similar question, writing:

> It is an interesting and unobvious assumption, which hasn't had the attention it deserves, that our responses to imaginary scenarios mirror our responses to real ones. .... that our intuition about what to do in the imaginary case is explained by the activation of the very mechanisms that would lead us to act in a real one...the responses activated through the process of imagining yourself to be in that situation look like the ones we expect would be activated if we really were in that situation. And, of course, there are moral accounts of long standing according to which the right thing to do is what a benevolent but fully informed observer would advise you to do: in which case we should perhaps put more stock, morally, in the questionnaire-answerer's counsel than in the switch-thrower's conduct.[103]

The later part of the quote, about the moral authority of a benevolent onlooker, is a problem for the idea that thought experiments can provide insight into moral questions,

---

103Appiah p.100-101

but is not uncontroversial.

The question of whether or not our reaction to hypothetical cases translates to real cases is raised by Goodenough, as a reason not to use thought experiments in bioethics. He writes, “And even if true beliefs are being generated, can we automatically assume that they will cover actual cases as well as the hypothetical case being posited?”[104] All of these concerns assume that there is some relevant difference between our understanding of hypothetical and actual cases. Indeed such a difference may exist between situations experienced first-hand and fictional or thought experimental narratives. If I give you two narratives describing situations in which a woman has to make choices about what kind of cancer treatment to pursue, one a real example and one a fictional example that is to the best of our current knowledge correct, people will likely not be able to distinguish the real from the fictional case if both are presented realistically. It would be easy enough to skew either the real case or the fictional one to make it seem fictional. If we are prepared to believe, first, that people cannot distinguish the fictional from the real case and, second, that intuitions that are generated in one case can cover other cases, then it may be argued that our intuitions about fictional cases will be no less applicable than those generated by actual cases. This brings up the question of the validity of the intuitions that are elicited by thought experiments, and by fictional or perhaps even actual but non-current scenarios. I believe that it is in part answered by Gendler's account of schemas, which locates our understanding of such hypotheticals in the same cognitive framework as experienced situations. I also discuss some related issues about using fiction in the next chapter, in the section on narratives. For now, the question at hand is the role of moral intuitions in

---

104Goodenough p. 9

thought experiments that are 'intuitions pumps'.

I Have this Niggling Intuition

In the book *Experiments in Ethics,* Kwame Anthony Appiah includes a chapter on intuitions. At the very beginning of this chapter he asks what the material of ethics is, and comes to the conclusion that it is intuitions that form the core of moral knowledge.[105] In so identifying intuitions as the basic building blocks of ethics the central question becomes the status of moral intuitions. Appiah spends a good part of his chapter on intuitions, looking at experiments that give reason to think that our intuitions are not reliable. Framing problems, making mistakes, and being influenced without being aware of the source of influence are all issues that Appiah raises. However, he also makes an interesting comment about the flaws that are found in our intuitions, reinforcing a basic idea that intuitions are at base right. This seems to me analogous to the way that the possibility of making addition errors does not show a basic problem with numbers, or even that we cannot do basic math correctly, just that we sometimes make mistakes. Appiah writes that "Learning that our intuitions are imperfect has, at least, one comforting implication: to be told that we sometimes get things wrong implies that we can, in principle, get them right."[106] He then balances that measured optimism with a more fundamental concern about what it is that our intuitions are drawing on: whether it is morality or social norms that are the source of 'moral' intuition.

It is clear that intuitions can be influenced by non-moral circumstances and elements. In reporting the results of a psychological experiment in which a choice between two

---

105Appiah p. 73
106Appiah p.116

identical but differently described scenarios give different results, it seems that if our intuitions differ, it cannot be because of any moral aspect, but is simply a cognitive error. But just as we are aware that our senses can be tricked by optical or auditory illusions, that we can make mistakes in our intuitions, especially when deciding in dilemma situations, is not a reason to give up on them entirely.

Bias in Intuitions

One problem with moral intuitions is the possibility of bias. Bias, it appears, is not limited to the obvious bias of self-serving interests, or even to more generally being situated as a particular individual with a specific set of experiences. “Stuck between Two Soup Cans: Limits of Rational Decision” by M. Champagne considers some psychological habits that influence us to choose between apparently indistinguishably identical options. Champagne reports psychological studies showing that right-handed people will describe the fluffiest towel to be the one farthest to their left, when the towels are in fact indistinguishable, that people can be influenced by the symmetry of a display, or will choose an object at a certain distance from the centre of their body. These psychological quirks have been exploited by marketers, who pay premiums to have products placed on shelves at certain heights in supermarkets, or at the centre of the section. What is troublesome is not that placement affects our choice, but that people are not generally aware of this bias, and so it seems that we may be influenced without being aware of the source of our decisions.

That our decisions are influenced by the dark recesses of our psyche, or the workings of our bodies, does not automatically implicate our moral intuitions. Appiah presents a

number of psychological experiments including a 2003 study Knobe describes two scenarios. In both a company CEO knowingly causes negative environmental effects. In one version the environmental impact is positive, in the other negative, with all other particulars remaining the same. The findings were that 82% of 78 study participants described the CEO's actions as intentional when there were negative effects, and 77% where there were positive effects.[107] The degree of blame or praise that participants attached to the CEO's action was also found to correlate to degree of attributed intentionality. Again, this looks like unconscious bias in our moral intuitions, and so is problematic for accounts that rely on moral intuitions.

In response to the Wolfenden report, a recommendation that male homosexuality be decriminalized in the context of the United Kingdom in 1957, Patrick Devlin argued that the law should be guided by the morality of the 'reasonable man' or the 'man on the Clapham omnibus'. The reasonable man is not a man who uses reason, but rather one who holds moral views that are commonly accepted.[108] Thus the reasonable man, if asked for his views, would express the moral intuitions of the nation. The way that this appeal to shared moral intuitions is phrased is jarring to modern nerves in a way, I believe, indicates that we are worried about moral intuitions, and for good reason. First and foremost, the idea that homosexuality should be illegal does not fit our expectations of most people in Canada. On further examination, the 'man on the Clapham omnibus' as used in British common law is clearly gendered, and moreover situated as a white, Christian, middle-class male. That we may not have shared intuitions, and that intuitions

---

[107] Knobe p. 193

[108] Aside from some moral relativist views, it is generally accepted that we can have shared moral intuitions that are not veridical.

are situated in time, place, gender, class, religion and race is a problem for calling upon such intuitions in the service of ethics.

Mental modeling offers both an explanation of how bias creeps into our intuitions and also a concern that bias is inextricably wound into thought experiments that elicit intuitions. Miščević's theory of chunked material that is fed into default hierarchies presents a picture in which our biases need not appear in the thought experiment, or in our explicit understanding of the models we use, but may sneak into our intuitions via the importation of large groupings of previous experience and background knowledge that is largely unexamined, though used to construct and wield mental models. Thus biases that exist in our background knowledge need not ever be considered or openly welded onto mental models to affect the results that mental models yield. By using large, inseparable chunks of experience, biases are given an open field to covertly enter into the modeling of thought experiments.

Having identified the problem of bias in intuitionism I will now move on to look at trolley problems as explanatory thought experiments, and Singer's drowning child as a motivational thought experiment. I will finish the section with an account by Hintikka that claims that doing multiple tests with variants on elements that we think may be morally relevant, such as the trolley problem variants, can serve to clarify what moral intuitions are.

Variability and Trolley Problems

Philipa Foot introduced the Trolley Problem in a discussion of the doctrine of double effect (1967). What has been taken up as an ethical thought experiment is the dilemma of

a trolley[109] driver who must choose to leave his runaway tram on its current course which will kill the five people on the track, or divert it onto another track which will kill the one person standing there.

> Suppose that a judge or magistrate is faced with rioters demanding that a culprit be found for a certain crime and threatening otherwise to take their own bloody revenge on a particular section of the community. The real culprit being unknown, the judge sees himself as able to prevent the bloodshed only by framing some innocent person and having him executed. ... To make the parallel as close as possible it may rather be supposed that he is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed. In the case of the riots the mob have five hostages, so that in both the exchange is supposed to be one man's life for the lives of five. The question is why we should say, without hesitation, that the driver should steer for the less occupied track, while most of us would be appalled at the idea that the innocent man could be framed. ... The doctrine of double effect offers us a way out of the difficulty, insisting that it is one thing to steer towards someone foreseeing that you will kill him and another to aim at his death as part of your plan.[110]

The moral intuition that is supposed to be at issue here is our reaction to 'unintended' consequences, and the use of trolley problems in moral philosophy is as clear a use of thought experiments as can be imagined. The original intention of Foot in constructing the trolley problem was, I believe, explanatory. It was used, alongside another thought experiment about a judge condemning an innocent man in order to promote the public good, to demonstrate applications of the doctrine of double effect in which one knowingly brings about actions that one could not or did not *directly* intend.[111]

Chapter Two looked at a number of thought experiments that purport to explain

---

109In light of the adopted and recognized name 'trolley problem' it is interesting to note that in the original article "The Problem of Abortion and the Doctrine of the Double Effect" the word trolley is never used.

110Foot p. 24

111Foot p. 23

freefall, however the differences between experiments were great enough that it makes no sense to say they are the same thought experiment. In contrast the trolley problem has bred variants in great profusion. Thomson suggests two of these. One is a much discussed version in which you are the subject who sees a runaway trolley and can save the five people on the tracks by pushing a fat man over a bridge and onto the tracks where his large body will stop the tram.[112] The fat-man variant has led in turn to scenarios in which the fat man is responsible for the runaway trolley, and where the fat man is standing on a looping tack, so that you can divert the tram to run him over, and because his body will block the trolley's movement, the trolley will not continue on the loop where it would kill the original five people on the tracks. The variations on the trolley problem highlight the differences in our moral intuitions by changing elements; they capture and examine and contrast our moral intuitions. I would classify these variations as explanatory, as the object seems to be to parse conceptual differences.

The website Philosophy Experiments[113] is a collection of online quizzes based on problems from philosophy. One of the quizzes is about the trolley problem, and begins by asking some 'preliminary questions' about the respondent's views on consequentialism. The respondent is then given the classic trolley problem, the fat-man variant, the fat-villain variant, and then a ticking time bomb-variant in which the fat villain is arrested and admits to having planted a nuclear bomb so that you must decide if it would be right to torture him to obtain the location of the bomb. The website claims to have obtained 102,101 responses as of July 2012, which break down as follows:

112Judith Jarvis Thomson, *Killing, Letting Die, and the Trolley Problem*, 59 The Monist 204-17 (1976)
113http://www.philosophyexperiments.com/

Overall Responses

	Yes	*No*
Should Casey Jones divert the train?	86%	14%
Should the fat man be pushed onto the track?	38%	62%
Should the saboteur be pushed onto the track?	75%	25%
Should the fat man be tortured?	75%	25%
Total responses:	102,101 [114]	

In the results from the Philosophy Problems website we can see, informally, a large sampling of the conflicts that psychologists have found in more controlled experiments. It is worth noting that there is significant variation in what people choose as the right answer for all of the trolley problem variants so far discussed. That framing issues affect these results is of course of concern when it comes to justifying the use of moral intuitions. However, what seems more problematic to me is that we do not have agreement on what our shared moral intuitions, if any, are for the trolley problem and for many other dilemmas. The take-home message of this section is that the moral intuitions elicited by thought experiments are variable, and in the following section I will present a theory that suggests a clarificatory role for such variability.

Well, You Won't be Justified, But...

If we temporarily set aside the question of how thought experiments can have

114http://www.philosophyexperiments.com/fatman/Default7.aspx

justificatory power, I believe that we can read an explanatory and motivational role for thought experiments in ethics into the account of thought experiments by Jaakko Hintikka. Even if thought experiments cannot justify, they may be able to clarify, communicate and provide incentive to act. In “The Emperor's New Intuitions,” Hintikka describes the use of philosophical intuitions as an artifact of the success of Chomsky's linguistic intuitions. He argues that the radical advances made in linguistics led philosophers to embrace the idea of intuitions. The problem being that the intuitions that Chomsky relied on were only about the usage of human language, whereas philosophers looked to intuitions not to understand the proper linguistic usage of the word ‘good’ , but for metaphysical and ethical facts about ‘good’. This led to problematic uses of intuitions by philosophers because in looking for truth, rather than correct linguistic usage as Chomsky does, philosophical intuitions run into the problem that moral realists face—namely that the concepts we hold do not necessarily have the force of natural laws, and so it is problematic to reach from how people use concepts to any underlying truth. In seeking knowledge beyond competence in using language or concepts correctly, it seems that philosophical intuitions must appeal to some kind of a priori truth. “Now, what conceivable theoretical rationale do contemporary philosophers' appeals to intuitions have? The embarrassing answer is: none... The vast majority of philosophical writers these days take the name 'intuition' in vain since they do not believe in Platonic anamnesis, Aristotelian forms, Cartesian innate ideas, or Kantian transcendental deductions”.[115] Without some good reason to think we have access to certain truths, there is no reason to treat our intuitions as having power to do anything beyond clarify

115 Hintikka p. 131

our use of concepts.

Hintikka does not, however, argue against the use of intuitions, or of thought experiments to elicit such intuitions. Indeed, he suggests that an experimental method will elicit more reliable intuitions. Earlier in this chapter we looked at a number of variations on the trolley problem. Using Hintikka's theory we can consider these variations as a means of seeing how, by changing the variables in the problem design, these variations might be used to delineate the factors that influence our intuitions. The same thought experiments will likely elicit a variety of responses, and by tracking the spread of responses we might be able to obtain a consilience of inductions by testing for overlap between sets and thus making something like a test of the parameters of the intuition.

Hintikka is concerned with generalizations from moral intuitions in a way that I find reminiscent of the problem of induction; the question is how to go from an intuition about one case to a general rule for moral conduct. For this reason Hintikka's description of performing series of thought experiments looks to me very similar to the reasons which are given to justify scientific experimentation as the most likely means of attaining knowledge. Hintikka writes

> I do not see any reason to deny that I can by means of such a thought experiment obtain objective knowledge about my own language and my own concepts, just as I can do so in the case of others by real experimentation. I do not need a concrete stage setting for the purpose. Thought experiments serve as well in the first-person case as real experiments in the third-person cases. If someone now wants to label the answers obtained by means of such thought experiments "intuitions," I do not have any objections, as long as it is realized what is involved.[116]

116 Hintikka p. 146

Hintikka argues that thought experiments are simply experiments that we perform on ourselves, which is not what I am proposing, but his idea of using multiple related thought experiments to delineate the boundaries of intuition is a useful one.

Types of Ethical Thought Experiments

In this section I differentiate explanatory, motivational and justificatory thought experiments with the aim of exonerating explanatory and motivational ethical thought experiments from the problems related to moral intuitions. I will briefly introduce the three categories, and give examples for each. The trolley problem is an explanatory thought experiment because it parses conceptual differences. The aim of the original trolley problem by Foot was to show that the unintended but deliberate causing of death is distinct from willing such a death. The trolley problem is an example that illustrates the fine distinction in Foot's idea of double effect. By contrasting this example with that of a judge who condemns an innocent man for the public good, and with the ideas about abortion which are the main thrust of the article, the trolley problem is used to separate out those intuitions we have which are specific to abortion, justice, and unintended consequences. The variants on the trolley problem can also be understood as working to separate the variables that influence our moral intuitions, in a way similar to the method that Hintikka suggests. This process has explanatory power, as it separates out the factors that affect our intutions in order to provide conceptual clarification. Another example of a explanatory thought experiment comes from Judith Jarvis Thomson.

In the famous example or thought experiment of Thomson's violinist, the reader is asked to imagine that they are kidnapped by the Society of Music lovers, and without

consent attached to a dying talented violinist. Upon waking up in hospital, the reader is then given the choice of staying attached to the violinist for nine months, and thereby saving his life, or of disconnecting the violinist thereby ending his life. Thomson used this to counter anti-abortion arguments which claim that if the fetus is a person, then abortion is wrong. If our moral intuitions in the case of the violinist, who is unequivocally a person, are that we have no moral duty to remain attached to him, then to end an unwanted pregnancy cannot be wrong in the event that a fetus is a person. However, our moral intuitions may not agree with Thomson`s, and one may suppose that if our intuitions are that there is a moral obligation to stay hooked up to the violinist, then the analogy falls flat. If Thomson's violinist is a thought experiment with explanatory power, then holding opposing intuitions is no problem for the thought experiment. What this thought experiment does is separate our ideas about gender and motherhood from the question of rights to bodily autonomy. By recasting the subject of the thought experiment not as a woman, or as a mother-to-be, the question of what obligation we have to save the life of another person is recast beyond the conventional, religious, or gendered conventions about what rights pregnant women have. By shearing the issues of gender and motherhood from the subject of the thought experiment Thomson clarifies the role that our background beliefs about women and mothers have in shaping our moral intuitions about abortion.

Thomson's violinist thought experiment is a success if the reader acknowledges that there are separate issues regarding who has rights to your body, whether motherhood has some strange ethical constraint, and whether there are moral duties specific to women's bodies and sexual morality. If these issues can be pulled apart, then the explanatory

purpose of the thought experiment has been fulfilled regardless of the moral intuitions elicited—whether or not the reader thinks it is morally permissible to unplug the violinist. If the reader creates a mental model without importing background knowledge (chunked material) about what pregnant women, separate from the class of people generally, may do, then Thomson's violinist has succeeded as an explanatory thought experiment. In explanatory thought experiments we learn by seeing what affects the models we construct, and how these models elicit different intuitions. Thus in the trolley-problem variants the models have different features, which then create differing intuitions so by comparing the elements used to create the models, and by seeing which features lead to the same or different intuitions the explanatory function is fulfilled. Thomson's violinist brings to light and challenges the background assumptions that are used in the abortion debate. By creating a model in which the standard chunked material cannot be automatically employed, Thomson's violinist is a thought experiment that has explanatory value.

If you saw a child drowning in a shallow pool, would you wade in to save them? What if doing so would ruin your expensive shoes? This is Singer's thought experiment from "Famine, Affluence, and Morality" which, I will argue, is intended to motivate. Singer very simply lays out his argument why those in affluent circumstances have a moral duty to help those who require it. He asks the readers whether they agree to two premises: that "Suffering and death from lack of food, shelter and medical care are bad"[117] and that "If it is in our power to prevent something bad from happening, without thereby sacrificing

---

117Singer p. 231

anything of comparable moral importance, then we ought, morally, to do it."[118] If the second premise is too daunting, he offers a weaker version in which we are not asked to sacrifice *anything* of moral value. Once the reader has agreed to these premises, Singer proposes that propinquity and unique ability make no difference to this moral duty, and thus that we have a moral duty to help people starving in Bengal because we can help them and their suffering is bad.

In a second year class on Global Justice[119] the students listened to the premises and generally agreed to them, but remained unconvinced. The questions were clustered around the practical requirements of ending world hunger, whether we have the capacity, and where responsibility for doing so should lie, namely, not on us. There were students who were convinced by the argument, but most were not. One said that he knew that something was wrong with it (the argument) but he did not know what. Having agreed with the premises, and having agreed that the conclusion results from the premises, people were still not convinced of the truth of the conclusion. When the drowning child thought experiment was introduced, heads started nodding. There seemed to be agreement that it would be morally obligatory to save the child at the cost of your clothes, and the questions shifted to whether distance can matter for moral duties. This is not to say that everyone was convinced and was motivated to give all their excess time and money to non-profit agencies assisting those in absolute poverty, but there did seem to be general agreement that saving the child was the right thing to do.[120] Singer's thought

---

118 Singer p. 231

119 Doan, Michael. "Famine, Affluence, and Morality." Justice in Global Perspective. Dalhousie. Dalhousie University, Halifax. 18 June 2012. Class lecture.

120 This change in directions can also be attributed to cognitive dissonance as people generally have the self concept as good people, and they might have been afflicted with the awareness that they would not

experiment motivates[121] by having the reader vividly imagine a scene in which their intuition is clear and strong. Thus it creates impetus by making a model that pulls in elements that lead to intuitions of a moral duty to help.

For all the criticism that it has received, Rawls's Veil of Ignorance[122] is one of the most widely known thought experiments. Rawls asks us to imagine ourselves in the original position, stripped of knowledge of our social, economic and historical position. Acting as free and equal agents, Rawls asks what social arrangements those behind the veil of ignorance would agree to. The strength of Rawls's Veil of Ignorance is precisely that it aims to re-frame experience, for the purpose of re-examining moral features. Making the hypothetical people in the original position ignorant of their characteristics, asks the reader to imagine herself stripped of identity, and given leave to agree on what institutions should form the basis of society. This process of re-framing and re-examining fits with Gendler's schematic mental modeling theory. Gendler describes a biblical narrative in which King David becomes involved in a relationship with a married woman, and then sends her soldier husband to battle with the intention of bringing about the husband's death. To show King David that his actions are unethical, Nathan presents King David with a story about sheep-stealing, which David identifies as unjust precisely because he fails to draw the connection between the story and his own case. By defamiliarizing the unethical acts that David performed, Nathan uses a story with morally

---

behave in the way that was morally right. (Thanks to my supervisor, Letitia Meynell, for raising this issue.)

121 It may be that Singer views his thought experiment as justificatory, (Thanks to my second reader, Kirstin Borgerson, for raising this issue.) however this is not made explicit in his writing, and the parallel use of arguments to make the point suggest to me that the thought experiment is not intended as justification.

122 Rawls, John. *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press, 1971.

relevant similarities to give David perspective on his own action.

Viewed in this light, moral and political philosophy have a secondary task that runs alongside the task of ascertaining what morality demands, namely, that of providing the reader with resources that enable him or her to make the perspective-shift that the moral stance requires at the moment of moral decision-making. In this regard, one of the tasks of such philosophical inquiry is to identify images that can play the role that Nathan's story did with respect to David: images that will bring the readers to reframe their experience of some morally valenced situation, in such a way that their apprehension of the morally relevant features of it are re-experienced in light of the scenario presented. It is this role, I want to suggest, that is played by some of the most famous thought experiments in moral and political theorizing.[123] Thus the idealized moral situation is treated as a schematic for the moral principles. If we take the advantages and disadvantages of intuitions seriously and are prepared to accept the potential for bias and for variability within individuals and across populations, mental modeling can also explain some of the means for which justificatory thought experiments are used. This is to say that the problems with intuitions pumps, that speak to whether or not thought experiments in ethics elicit anything more than transient and biased intuitions, continue to pose potential problems. Social contract theories rely on thought experimentation for their justification. The reason for accepting the principles advanced by Rawl's theory of justice is simply that our imagined actors behind the veil of ignorance would accept such structures. Thus the justification is the thought-experimental results.

123Gendler p. 83

Mediative, Conjectural, Direct and Destructive Thought Experiments

In Chapter One I introduced Jim Brown's taxonomy of thought experiments, which divides thought experiments into destructive and constructive, and then further divides constructive thought experiments into the categories of mediative, conjectural, and direct. Having introduced in this chapter a distinction among justificatory, explanatory and motivational thought experiments, I will now apply Brown's categories to ethical thought experiments, and look for overlap with my categories of justificatory, explanatory and motivational thought experiments.

It is tempting to see many ethical thought experiments as destructive simply because many claims run counter to some theory in ethics. Foot's scenario of a judge condemning an innocent man in order to ensure the peace is not a destructive thought experiment, because the purpose of it is to demonstrate the doctrine of double effect. The same thought experiment would, however, be classified as destructive if it was used as an argument against utilitarianism, to show that utilitarianism is incompatible with justice. Thus while many ethical thought experiments will oppose some other assertion or theory, a diversity of theories is accepted in ethics, and any interesting thought experiment is likely to run afoul of some of them.[124] Thus only those which make primarily a negative rather than positive claim should be treated as destructive thought experiments.

Mediative thought experiments, I believe, share characteristics with both explanatory and motivational thought experiments. Mediative thought experiments demonstrate a problematic aspect of an articulated theory, and work by either (or also) clarifying or

---

124Thanks to my supervisor, Letitia Meynell, for raising this issue.

providing support.[125] Where the purpose of the thought experiment is to shed light on the theory, I would liken it to an explanatory thought experiment, and where the purpose is to strengthen a position, I suggest it may be likened to a motivational thought experiment. I liken conceptual clarification, which is one of the purposes Brown assigns to mediative thought experiments, to explanatory power. I do this because I have treated thought experiments such as Thomson's violinist which delineate and bring into focus ethical ideas as explanatory, and this function is very much like the pulling apart of ideas involved in conceptual clarification. That providing support for a theory is like motivating an ethical imperative, idea, or belief is perhaps less convincing. Thus the link between motivational and mediative thought experiments may be weaker.

Direct thought experiments are justificatory, as the goal of a direct thought experiment is to establish a clear result, and from this result support a theory.[126] The work to be done is in taking these results and getting to a theory which fits the results. This is, I would say, the way that justificatory thought experiments are supposed to work. A moral intuition is elicited that is so clearly true that theory is then built to accommodate it.

Conjectural thought experiments propose a thought-experimental result from which theory is developed. The result of the thought experiment is open to contest, and a theory is called upon to explain it.[127] In this case it is the theory that must do the heavy lifting, as the results are contested. This is how justificatory thought experiments are more likely to play out; the moral intuition is contested, and a theory is brought in to give principled support. In seeing the overlap that is possible between Brown's categories of scientific

---

125Brown p.36
126Brown p.41
127Brown p.40

thought experiments and the function of ethical thought experiments we see that not only can the content of both ethical and scientific thought experiments be explained using mental modeling, but that the distinct differences between ethical and scientific thought experiments in function do allow for both to play many of the same roles, though in distinct ways. These overlapping but not identical functions will come into play in bioethical thought experiments which pull from the traditions both of scientific and ethical thought experiments. The difference between scientific thought experiments which work as knowledge claims and ethical thought experiments which act as intuition pumps is pronounced. However, Brown's taxonomy can also throw light on the function of ethical thought experiments.

Ethical Thought Experiments in Sum

In this chapter I have argued that ethical thought experiments elicit our moral intuitions and make concrete our moral commitments. The latter purpose is especially relevant for bioethics and for other applied ethical disciplines because thought experiments make the move from abstraction to discrete instances. Thought experiments as ethical 'intuition pumps' face problems. However these problems largely stem from a concern with the status of moral intuitions that is not confined to thought experiments, but is a general metaphysical problem for ethics. The problems that moral intuitions face are bias, and variation both within the intuitions elicited in individuals by different thought experiments such as variants on the trolley problem, and variation in the intuitions that thought experiments evoke in different people, and in different populations. However, thought experiments hold out the promise of elucidating what moral intuitions are held and evoked, which is of benefit in ethics and other philosophical disciplines that

rely on intuitions. Thus that intuitions are variable is not necessarily a problem, but is instead a source of information about intuitions.

By dividing thought experiments into motivational, like Singer's drowning child, explanatory like the various Trolley Problems, and justificatory, I argued that the problem of moral intuitions only troubles justificatory thought experiments. This means of dividing up ethical thought experiments can be fitted with Brown's taxonomy of scientific thought experiments to contrast scientific and ethical thought experiments, to see that there are differences in content but significant similarities in purpose. In addition mental modeling is, I argue, the best account of the working of thought experiments in ethics. Having considered ethical and scientific thought experiments, I will now turn to bioethical thought experiments, which include elements of both science and ethics.

Chapter 5 Bioethics

Introduction

In the introduction I offered as a foil Goodenough's article objecting to the use of thought experiments in bioethics. Goodenough is especially leery of ethical thought experiments, though he is forced to allow some place for thought experiments even in the realm of morality. In questioning whether or not the intuitions that are elicited using thought experiments are good ones Goodenough points to the use of thought experiments in ethics. "Such cases must be part of the ethical training of any medical professional or related discipline. And here hypothetical cases can often be substituted for by real cases"[128] The key words in the sentence are 'can' and 'often'. By saying that it is often possible, an implicit space is created in which real cases cannot be substituted for hypothetical ones. In this chapter I will respond by presenting three arguments for the usefulness of thought experiments in bioethics. The first is that thought experiments will be better than real examples when they more clearly demonstrate a point than a case study can. The second argument is that bioethics deals with new technologies, and it is advantageous to consider the ethics of a new technology before it is put into practice, which means that real cases will simply not be available, and thus thought experiments are by default the only option. The third argument is that in ethics there will be case studies that are unavailable because it would be unethical to enact them.

In issues of bioethics that consider emerging or new technologies, it is unlikely that there will be any real cases available antecedent to ethical consideration. There will also

---

128Goodenough p. 8

be situations that, while not involving a new technology, are novel is some other ways, or have not been recorded and so are unavailable as case studies. We may also ask if there are situations that are, by their very nature, impossible to replace with real cases. Thus it may be argued that there is a range of cases that cannot in fact be found in real life, and so a claim might be made that where real cases are unavailable or impossible, hypothetical cases are allowable. But are real cases always to be preferred to hypothetical ones? Goodenough does not entirely denounce thought experiments, ceding them some place in science at least, but in general his view is that hypothetical cases are always second-rate, that it is always better to perform the experiment in a laboratory rather than the 'laboratory of the mind', and that when it comes to philosophical thought experiments there may be some slight place for them is showing contradictions, but that anything involving intuitions is suspect.

Bioethics is a field of study that is broadly concerned with how the science and practice of medicine and technologies affect bodies and what the ethical concerns with and implications of this may be. As such science and ethics come together in this field in a unique way. Thus the question of how scientific and ethical thought experiments are the same and different must be addressed. The *Stanford Encyclopedia* entry on Feminist Bioethics written by Anne Donchin says the following about the bioethics movement:

> The bioethics movement was triggered by protest against such gross abuses of medical authority as the Nazi doctors' experiments on unconsenting concentration camp inmates and the Tuskegee Syphilis Study, a forty year “experiment” on poor black men who were misled into believing they were receiving therapy. ...It has generated a massive literature ranging over a broad array of moral problems that arise within biomedical research, the health care professions, and the institutions that deliver health care services. Its reach extends to philosophical and

> legal issues from the beginning to the end of human life, to areas of biology and genetics on which medicine draws, and to research that seeks to expand the knowledge base of medicine.[129]

The cognitive upshots of ethical and scientific thought experiments can be explained by mental modeling. Scientific thought experiments look to the results of the thought experiments—the conclusion is drawn from what we imagine happens. Brown writes that, "We would agree (as would most people) that a real experiment carries us from a perception (and some possible background propositions) to a proposition (a statement of the result). The so-called experimental result may be the culmination of a great deal of theorizing and calculating, but somewhere along the way the experimenter has had to look at something.... I hold that thought experiment has a similar structure."[130] In Galileo's falling objects thought experiment, it is imagining the moment of impact that informs us that the weight of falling objects cannot affect the speed with which they fall. For ethical thought experiments the cognitive upshot is derived not from the conclusion, which is often a simple yes/no answer, but rather from the principles and intuitions that inform the conclusion. It is answering why, not answering the question itself that is informative. Thus ethical thought experiments may be seen as explanatory, where what is being clarified is moral commitments and intuitions. In bioethics the scientific and the moral intersect, in answering questions about what it is right to do in the realm of the applied human sciences. In the next chapter I will argue that science fiction draws together the scientific and the ethical in thought experiments that are particularly suited to exploit the cognitive upshots of bioethics. In this section my goal is to argue that thought experiments in bioethics bring together the purposes of scientific and ethical

---

129Donchin n.p.
130Brown (2004) p.35

thought experiments, and are of value.

Exemplum

Bioethics is rife with case studies. In a way that philosophy of science and ethics are not, real cases are used with abandon in bioethics textbooks and articles. There is a sense that for an applied discipline it is the real cases that matter. Thus the use of examples is well accepted in bioethics. The work for thought experiments is in showing that examples that are hypothetical are of worth. I suggest that we imagine a spectrum of examples, with fictions explicitly not related to any true states of affairs standing at one end, and extremely well documented and detailed cases studies on the other. Thought experiments fall between complete fiction and case studies, and I will argue that thought experiments can be informative not in spite of not being real cases, but because they are not real cases.

Interestingly, the line between case studies and fiction has been challenged. Freud's case studies are perhaps the most widely read, and are now acknowledged to be as much fiction as fact. We can of course simply dismiss this as lazy science, or slack reporting, but the fact remains that Freud's case studies are gripping, memorable and still a staple of psychology classes. Case studies are of course a form of narrative, though not purportedly fictional narrative. In selecting the information to report, and the details to include or exclude, and how to present the information, a narrative is constructed.

Both case studies and thought experiments work as examples, and so have some of the same purposes. That examples are used in bioethics therefore does half the work, and if it can be shown that sometimes thought experiments make better examples, then in those

cases thought experiments will be the better choice. Thus the question is not whether real cases or hypothetical ones are useful in bioethics, but what makes a good example and whether a thought experiment can have these relevant features.[131] By analogy, it is easier to identify birds using well-drawn field guides than photos of the birds, because expert illustrators who know what features distinguish species can ensure that all such features are visible and prominent. In the same way a clear and directed fictional account may be more informative, and thought experiments are positioned to act in this way. Real cases are complicated, and it may be unclear what is and is not relevant to the issue at hand. Thought experiments have purposes other than providing empirical evidence, and can be used to serve these purposes in the field of bioethics.

The use of first-person narratives is one example of how thought experiments can be better than real examples. To return to Thomson's violinist thought experiment, one element of this thought experiment that cannot be captured by a case study is that of engaging the reader as the subject of the thought experiment. The choice is not one that you make on behalf of another specific or generic person, but rather a choice that is imagined happening to you. Making you the subject of the dilemma is simply impossible in a case study. Thomson excises paternalism, and by casting the subject as *you* she also removes issues of gender, race, class, religion and other social signifiers that she wishes to assert are irrelevant. A case study, unless it is excised of all the information that makes it a concrete example, will contain some identifying information. Thus the first-person narrative has, for Thomson's violinist thought experiment, two clarificatory functions that cannot be created or found in any case study. This is one way that thought

131Thanks to my supervisor, Letitia Meynell, for raising this issue.

experiments may be better than real examples.

Technical Difficulties: Please Stand By

The second reason for using thought experiments in bioethics is that technical impossibility is a particular problem for bioethics. There are of course issues such as euthanasia and abortion for which there is a wealth of studies, case studies, and real examples. However, bioethics is also concerned with speculative issues such as human genetic engineering for which there is no specific precedent. This is not to say that debates over new issues will occur in a vacuum. Genetic engineering raises concerns about eugenics, which is an issue with a long history. It also raises concerns about using a new technology for 'bad' ends, a debate reminiscent of mid-century concerns about nuclear technology.[132] The tradition of narrative case studies in bioethics has indeed the effect of opening the field for thought experiments. General principles and arguments are open to bioethics, but in a field that has made such good use of case studies it makes sense to make use of thought experiments. By having some of the concrete, causation-related, clarifying features of case studies, thought experiments fill a niche in bioethics. In emerging and speculative areas of bioethics there are no case studies to be had. Thought experiments are thus uniquely suited to fill this niche.

The system of regulations that is widely used makes it difficult to have a new human technology, such as a drug or medical procedure, approved, and once such a product is in use it is difficult to have the applications of the technology changed. Therefore, once a new technology is on the market, it becomes more difficult to change the regulations.

---

132 Thanks to my second reader, Kirstin Borgerson, for this idea and this example.

Regulatory approval for drugs and devices can be withdrawn if evidence of harm, such as drugs with serious negative side effects, is found.[133] This evidence to change regulations may be hard to come by due to a lack of controls in the general population, thus it may be more difficult, though not impossible, to withdraw devices and drugs on the are on the market.[134] Looking ahead to the effects such a technology will or may have is therefore of great interest in bioethics, as to effectively regulate, it is virtually incumbent to look at future, rather than current, technologies.

One function of thought experiments is to act as precursors to real experiments. Galileo's falling-bodies thought experiment was a real experiment waiting for the availability of a vacuum in which to be performed. There are many more experiments in science that we lack the technology or the resources to perform now. In contrast ethical thought experiments are often experiments that we do not want to conduct, which brings us to the third reason.

On the Island of Dr. Moreau

The perfect case study will not always occur organically, and in many cases, particularly ones with moral dimensions, it would be wrong to enact the scenario of interest. Thomson's violinist combines the current technical impossibility with an experiment that we would not want to actually run. There are at present no therapies that require one human being be hooked up to another for a term of nine months, but even if we could use such a method to cure cancer, experimenters should not kidnap a subject in

---

133Thanks to my second reader, Kirstin Borgerson

134Thanks to my second reader, Kirstin Borgerson

order to find out what they would decide to do and why. Philosophical and literary utopias and dystopias[135] also share this characteristic of being both being infeasible and, at least in the case of dystopias, being experiments that we don't want to have happen. One example can be found in H. G. Wells' novel *The Island of Dr. Moreau,* in which a shipwrecked scientist finds himself on an island belonging to a discredited vivisectionist, Dr. Moreau, who is engaged in experiments in creating human-animal hybrids. *The Island of Dr. Moreau* depicts experimentation that may be ethically interesting but cannot be performed. In such cases real cases might be informative but for ethical reasons cannot be performed.

Conclusion

I have argued that thought experiments are useful in bioethics. While it may be argued that where real cases exist they are better, my first argument, that thought experiments can identify and clarify as case studies may not be able to, does not rest on a dearth of real examples. Instead I identify situations where thought experiments will be more effective. In addition mental modeling theories give us reason to think that thought experiments allow for forms of reasoning based on the use of mental models that are not available to us via argumentation. Given the close connection of narratives to mental models, case studies and thought experiments both allow access to non-propositional, model-based cognition, which is distinct from the thinking that argument alone produces. This does not give us reason to prefer thought experiments to real cases, and indeed real cases are certainly useful in bioethics and elsewhere, and I am by no means advocating

---

135 This may be the case particularly in "negative thought experiments where the point is that some course of action should not be pursued" (Thanks to my supervisor, Letitia Meynell, for raising this issue.)

that we shun real examples as a class in favour of thought experiments. However, sometimes thought experiments will be better, and sometimes real cases will be unavailable. Given the cognitive advantages that access to mental models provides, there is good reason to include thought experiments in bioethics.

Chapter 6 Narrative

Introduction

"We novelists are used to toying with grand philosophical theories and we are certainly not apt to clarify them: we falsify them but we *humanize* them"[136]

Grazia Deledda, a Nobel Prize winner for Literature, thus expresses her view of the interaction between literature and philosophy. This quote fits with the idea that thought experiments give concrete expression to ideas, which may have the effect of changing our focus and so our intuitions and understanding. However, Deledda, an author known for the use of philosophical ideas in her writing, denies that this 'humanization', or change of scale, brings ideas into focus. She writes that the act of humanizing, which I would judge akin to the move from pure theory to applying ideas within imaginary worlds, falsifies rather than clarifies.

In this section on narrative I will argue that fiction can act as, and be, thought-experimental. First, I will look at a number of objections to the idea of literary thought experiments; then I will examine a family resemblance theory that holds the promise of discriminating between fiction that is, and is not, thought-experimental. The obvious choice for literary thought experiments is in the field of ethics, given the vast body of art that is primarily concerned with morality. For the role of ethics I will look at Carroll's article about virtue ethics in literature . However, it is not only ethics that is explored through literature; Davenport's article takes scientific thought experiments as a starting point. I argue that philosophy can usefully import scenarios from fiction to use as thought

136As translated by Margherita Caput, p. 16, italics added for emphasis

experiments, and also that some authors of fiction engage in thought experiments. In the following chapter I will argue that science fiction works that are primarily focused on 'what if' questions are thought-experimental.

Fiction is About Entertainment, not Ideas

The foremost objection to the idea of literary thought experiments is that however fictional works may be interpreted, borrowed or adapted by philosophers, the authors of the works did not intend to create thought experiments. Indeed, if the purpose of fiction is to entertain and not to explore ideas, it seems the best we can do is find thought experiments in literature, and works of literature themselves can never be thought experiments. In this section I look at an article arguing that George Eliot intended her novels to be philosophical discourses and that they succeed as such.

In "The Art and Philosophy of George Eliot", Moira Gatens argues that the novels of George Eliot are philosophical and literary works, and must be understood as both philosophy and literature: “her novels should be understood as attempts to practice philosophy in an alternative key. Her decision to write novels rather than conventional philosophy reflects her desire to actively engage the imaginative and affective, as well as the cognitive, powers of her readers.”[137] Eliot described her own works as “simply a set of experiments in life—”[138] and Gatens identifies this as a recurring motif in Eliot's work and in Eliot's own writing about her work.

Imagination is one of the subjects of some of Eliot's major works including

---

137Gatens p. 76
138As quote din Gatens p. 81

Middlemarch. In *Middlemarch* Eliot positions the imagination as a motive force in the life of the protagonist, Dorothea, who is led into a problematic marriage by the power of her moral imagination. Gatens considers the translations from German to English that Eliot made of Feuerbach's *The Essence of Christianity* and Spinoza's *Ethics*, and argues that these translations heavily inform the philosophical nature of later novels by Eliot. Both Spinoza and Feuerbach were concerned with the role of imagination, and with the idea of moral knowledge as a combined power of imagination and reason. Gatens writes that "Both Spinoza and Feuerbach argued that a primary task of philosophy is to articulate the relation between imagination and reason. Understood and used correctly the imagination is not an epistemological defect but rather an essential power of human beings... Adopting aspects of Spinoza's and Feuerbach's philosophies, Eliot treats the imagination as a source of knowledge."[139] What makes Eliot singular is that in using the medium of fiction she wrote philosophy and literature together in a *joint* work that harnesses the imagination and offers a discourse about it.

It is not a great leap to say that Eliot's 'experiments of life' are thought experiments, and that her project of combining the strengths of imaginative and cognitive reasoning or argumentative knowledge is a plausible account of how literary thought experiments work. In the following section I will present an article that argues that literary thought experiments are possible and which takes *Middlemarch* as an example. Eliot, who was unquestionably an author of fiction, wrote with the intention of exploring philosophical ideas, and her works succeeded in doing so by creating worlds in which the concepts she explores could be imagined.

139Gatens p. 80

Scientific Strength, Didactic Writing

I argued that bioethical thought experiments may be characterized as a joining of elements from scientific and ethical thought experiments. In the following chapter I will argue that it is precisely these thought experiments for which science fiction is particularly useful. I further seek to reconcile scientific and ethical thought experiments in the medium of science fiction, as a genre of literature. In "Literature as Thought Experiment (On Aiding and Abetting the Muse)" Edward Davenport denies that art is merely emotional and non-cognitive. He claims that thought experiments bridge the gap and oppose the idea of a polarity between art and ideas.

Davenport writes, "... literature presents a plausible world—a world which makes a good framework for testing out our own ideas as well as for testing out the ideas presented to us in the novel."[140] According to Davenport, these plausible worlds are used to test ideas, not only by philosophers but also by readers in general. This is a particularly interesting account, because it expands the definition of thought experiments in literature beyond those instances when philosophers take a scenario from fiction and *call it* a thought experiment. This expansion fits Davenport's definition of thought experiments. For him, "Thought experiment means testing hypotheses in the mind-- logically rather than physically. This may be done by making deductions from the hypothesis to see what must follow if it is true."[141] This experimentalist definition is fitted to the idea of plausible worlds in fiction by Davenport, who talks about thought experiments as worlds in which philosophical problems are acted out, and as an act of scientific experimentation

140Davenport p. 298
141Davenport p. 281

that is done internally. This idea of testing ideas in the mind is also entirely compatible with a mental-modeling account of how thought experiments work. The idea of testing hypotheses has notable similarities with the idea of creating mental models, because to test these possibilities in our minds we must build the scaffold on which to test them, and bring in the background knowledge to make our tests accurate and informative. This may be seen in the way that Davenport talks about political thought experiments: "Good thought experiments are expected to be informative both conceptually and empirically. With Marx's *Das Kapital*, or Freud's *Civilization and Its Discontents*, or Weber's *The Protestant Ethic and the Spirit of Capitalism*, few readers would accept the proposed model of society only because they agreed with its logical-conceptual implications. Fewer still would accept such a thought experiment just because they believed that it had passed one or another physical-empirical test."[142] Consistent with the mental modeling account, Davenport treats thought experiments as created worlds in which authors and readers can test hypotheses. He writes, "Some authors catch our attention by making certain problems important for us which had not been so before... literature presents a plausible world—a world which makes a good framework for testing out our own ideas as well as for testing out the ideas presented to us in the novel."[143]

In Chapter One I introduced Davenport as a proponent of the idea of stand-alone thought experiments in science. This reliance on thought experiments that stand on their own, without the need for an actual experiment as confirmation, is essential to Davenport's attempt to root literary thought experiments is the scientific tradition. Free-

---

142Davenport p. 283
143Davenport p. 297-8

standing thought experiments bridge the purported difference between the conceptual and the actual, and the use of free-standing scientific thought experiments has the effect of opening up the question of what we learn through fiction. Davenport uses Eliot's *Middlemarch* as a thought experiment about marriage in the cultural context of England in 1830 as experienced by rural gentry. Davenport writes that, "Dorothea's ideas about marriage are grounded in the assumption that self-renunciation and regard for others is the key not only to a moral life but to a happy life. This is a general theory about the individual in society which George Eliot enables readers to test, by giving us the case of Dorothea's own actions in accord with this theory. This literary thought experiment enables us to examine both evidence which seems to confirm such a theory, and evidence which seems to criticize it."[144] The ideas are sociological, and in addition to being examined and referenced by social scientists, *Middlemarch* was written by an author who left correspondence indicating that the exploration of ideas was one of her reasons for writing the novel. Moreover, *Middlemarch* is a novel which creates a plausible scenario for testing the proposed ideal of marriage.

What is interesting about Davenport's account is that authors, readers and the literary works are said to benefit from the process of thought-experimental testing. In writing, authors run their own mental experiments to check both whether their fiction is internally and externally consistent, and also whether it is plausible and believable;[145] one could say that in constructing a mental model of the proposed world, authors check how the parts fit and work together, and how the background assumptions interact in the model. Thus

144Davenport p. 302
145Davenport p. 292

the thought experiment is an intrinsic part of the work, and far from taking away from the merit of the fiction by making it didactic and thus not artistic, thought-experimental narrative artworks works may be the stronger as artworks for being thought-experimental.

Davenport considers the use of thought experiments in literature as a means of improving the literature. He argues that literature is about ideas, and thus better ideas will make for better fiction, which links philosophical and literary development. Davenport makes the sweeping claim that it is simply evident that literature acts as thought experiment, and also claims that the literary works that most clearly function as thought experiments are marginalized within the canon of literature:

> The idea that literature can be viewed as thought experiment is at once very commonsense and very paradoxical. It is commonsense because it is easy to see that many kinds of stories are like thought experiments, including Aesop's Fables, the parables of Jesus, the dialogues of Plato, the speeches of Thucydides' *Peloponnesian War*, Thomas More's *Utopia*, Swift's *Gulliver's Travels*, Johnson's *Rasselas*, Mary Shelley's *Frankenstein*, Wells's *The Sleeper Awakes*, Orwell's *1984*, and Asimov's *Foundation Trilogy*. All these stories dramatize certain hypotheses about society and enable us to see the logical conceptual implications of the hypotheses, and so it seems commonsense to say that such stories are thought experiments. ...[and yet] So paradoxical does it seem to view literature as thought experiment from which we can gain cognitive knowledge, that literary critics and theorists have frequently classed such books as I mention above as 'didactic' and therefore 'outside of' or 'on the margins of' literature. This is done to indicate that the more a work of literature is like a thought experiment, the less it is like literature (because the less noncognitive and non rational it is)[146]

The marginalization of such 'didactic' literature, for the reason that it is aimed at making a point or championing a view, is in conflict with the influence of the writers who saw criticizing ideas and imparting knowledge as one of the goals of literary works. Davenport attributes this view that literature is about ideas to Voltaire, Rousseau, Diderot,

---

146Davenport p. 283-4

Dickens, Eliot, Dostoevsky, Tolstoy, Shaw, and Brecht.[147] And given that these authors are among the most respected literary figures, it seems that there is a burden of proof on the critic of didactic literature.[148] Given that some of the greatest writers see value in justifying views through literature, there must be a strong argument against didactic literature being literature. For later consideration I point out that the list of didactic works is made up almost entirely of utopias, dystopias and literature that could be classed as science fiction: Thomas Mores's *Utopia*, Wells's *The Sleeper Awakes* and Orwell's *1984* being dystopias, and both Orwell's *1984 and* Wells's *The Sleeper Awakes,* as well as Swift's *Gulliver's Travels*, Mary Shelley's *Frankenstein*, and Asimov's *Foundation Trilogy* arguably fitting under the umbrella of science fiction.

The view that literature can be primarily about ideas and that author and audience both use the fictional world to test ideas is necessary to claim the existence of literary thought experiments. However, there are objections to this, and in the next section I will respond to some objections against literature as a thought experiment.

Some Objections

That literary thought experiments have ethical features and cognitive upshots is an important claim for my thesis, and in this section I will generalize from a specific claim about the existence of virtue-ethics thought experiments in literature to include ethics more broadly. In “The Wheel of Virtue: Art, Literature, and Moral Knowledge,” Noel Carroll gives a set of reasons to consider some literature to be thought-experimental and responds to arguments against treating literature as thought experiments. The objections

147Davenport p. 289
148Thanks to my supervisor, Letitia Meynell, for raising this issue.

are 1) the banality argument, which says that literature does not give us any new ideas, 2) the no-evidence argument, which points out that thought experiments are not real experiments and so do not provide evidence, and 3) the no-argument argument, which charges literature with a failure to make arguments which can then be debated and analyzed. These three arguments are not only arrayed against the belief that literature can act as thought experiments, but are also used more generally to deny that literature can be a source of knowledge.[149]

The response that Carroll makes is specific to virtue ethics, which is not of special concern to my thesis. I will therefore restrict my interest to Carroll's claims for thought experiments in literature more generally. In addition to eschewing Carroll's focus on virtue ethics, I do not share his view that thought experiments serve only as a means of conceptual analysis. The mental-modeling view of thought experiments is not inconsistent with conceptual analysis but allows that thought experiments can be used for other tasks. Thus I take Carroll as a starting point, and from his defence of virtue ethics and thought experiments as conceptual analysis I will argue that ethical thought experiments generally may be found in literature, and that these are best understood using mental modeling.

The first reason that Carroll gives to think that there are literary thought experiments is that philosophers borrow from fiction to propose thought experiments. Carroll offers the following example: “Encountering the Socratic doctrine that a person who knows the good cannot choose to do evil, the philosopher may respond by drawing attention to the literary cases of Milton's Satan, who declares, “Evil be thou my good,” as well as

149Carroll p.7

Shakespeare's Iago and Melville's Claggart. The reader, using her conception of what is humanly possible, recognizes that such personality types could obtain".[150] That thought experiments may be borrowed from literature does not, however, mean that literature can be a thought experiment on its own and without the mediation of a philosophical text. The reason that Carroll gives for treating literature itself as a thought experiment is that it can perform the functions that thought experiments perform. The defence of this claim comes in responding to objections that charge that literary thought experiments do not perform useful functions.

The first such objection is the banality objection, which says that literature does not give us any new ideas. Carroll's response to the banality objection is that in virtue ethics there is learning that has to do not with the wide swathes of right and wrong, but with being attentive, picking out the correct moral cues, and knowing how to best apply ethical principles in a given situation. This is not necessarily new information, and may indeed involve a message as seemingly trite as 'don't be cruel'. However, there is knowledge and learning involved. Carroll summarizes this position, saying that, "Since the education involved concerns the refinement of our grasp of virtue concepts, it is not best described as banal or platitudinous, but rather as affording added insight into that which we already know."[151] In applying this defence to ethics more generally I think that there is room for the ideas of attentiveness that care ethicists including Joan Tronto have raised. The process of recognizing how to care and what the appropriate actions of care are seems to me to be another instance in which learning does not necessarily require the

150Carroll p.9
151Carroll p. 11

new and unexpected, but may be found in examples which clarify known concepts. Indeed, the questions of when and how to apply principles is a staple of the field of applied ethics. That information may be conveyed that is not new to the writer or the reader, but may clarify, illuminate, or even motivate moral actions, I believe, is a good response to the banality objection beyond the context of virtue ethics.

Carroll also allows a role for emotion in identifying moral features, in what we attend to and how we react, saying that "... emotional responses are part of the mix of factors that are engaged in deliberating about the application of virtue concepts in reaction to fictional thought experiments."[152] This is, for Carroll, a response to the possible objection that reading literature as thought experiments fails to account for the important role of emotion in appreciating or experiencing art. However, I believe that the role of emotions in literature is also a response to the banality argument. Just as care ethics suggests that learning to care is a process of recognition and practice and so envisioning different scenarios with the same underlying, familiar principles is not useless if we allow emotions a place in morality, so too will the emotional responses we have to fictions with the same familiar principles be informative. By creating an emotional as well as an intellectual response to moral issues, thought experiments may be ethically useful. This is related to Nussbaum's ideas that by informing and forming our emotions, literature can result in moral development.[153]

The perhaps surprising conclusion that may be drawn from these arguments is that literary thought experiments may be more effective than philosophical thought

---

152 Carroll p.18

153 Nussbaum, Martha. *Upheavals of Thought: The Intelligence of Emotions*, (2001) Cambridge: Cambridge University Press.

experiments for ethics. Many ethical thought experiments create an emotional response. The idea of choosing to push the fat man onto trolley tracks also evokes emotion. However, it is more likely that people will become emotionally invested in a literary work, and the context-rich, character-driven, situated form of literary works is a richer source of moral emotions than a short philosophical thought experiment can be. If we take the idea of moral emotions such as resentment seriously, or even if we simply accept that emotion is an integral part of understanding and acting morally, we may have a reason to prefer literary thought experiments. Carroll writes that

> It is often said that literary examples are far more effective in eliciting ethical understanding than are abstract philosophical arguments. One reason for this is that, though more simplified and structured than actual cases, they are much richer in detail—about motives, feelings, circumstances, social relations, and interconnected personality traits—than typical philosophical arguments and thought experiments... Thus, the elaborateness of literary examples is not grounds for disqualifying them as thought experiments, but rather grounds for appreciating them as thought experiments that have special cognitive requirements and advantages.[154]

I will return to the idea that literature is particularly adept at capturing ethics in great richness and depth, and that literary forms may, by their very nature, lend themselves to ethical understanding. Revisiting the quotation with which I began this chapter, by humanizing theories--by bringing them to human scale and by including the rich detail of human lives—novelists clarify the ethical aspects of human lives.

The recurring motifs of literature, be they boy-meets-girl, villain-comes-to-a-bad-end, or journey-of-discovery, will be banal in one sense. But as I argued above, the nuances may be of value, and there will always be differences in nuance. I would also argue that

---

154Carroll p.19

the objection to fiction as banal does not apply to literary thought experiments that present novel situations, ideas, or technological applications. In the case of science fiction as a bioethical thought experiment, we often have new ideas presented in literary form. The story of *Neuromancer* by William Gibson could be described as merely a boy-meets-girl story, but to do so would miss the new ideas presented in this novel. In *How we became posthuman: virtual bodies in cybernetics, literature, and informatics,* Katherine Hayles credits Gibson's cyberpunk oeuvre as one of the forces that shaped computer technology. She writes, "I want to resist the idea that influence flows from science into literature. The cross-currents are considerably more complex than a one-way model of influence would allow. In the *Neuromancer* trilogy, for example, William Gibson's vision of cyberspace had a considerable effect on the development of three-dimensional virtual reality imaging software."[155] There are instances of science fiction that envisions the results of a new technology for society in advance of the technology`s being released or generally adopted.

The second objection is the banality objection, which points out that thought experiments are not real experiments and so do not provide evidence. Carroll's response to the no-argument argument recapitulates John Norton's position that thought experiments are arguments. Carroll describes them as 'instruments of argumentation'[156] rather than as arguments, and as "function[ing] as arguments."[157] However, the difference is a slender one. Carroll defends the view of literature as thought experiment, and from there claims that because they have comparable structural features *and produce*

155Hayles p. 21
156Carroll p.15
157Carroll p.7

knowledge, thought experiments are sufficiently like arguments that the no-argument argument fails to hold water. I am very sympathetic to the arguments that literature can possess structures and functions that are thought-experimental. However, I do not agree that thought experiments are arguments. In Chapter Three on mental modeling I argued that thought experiments are experientially, perceptually, and functionally different from arguments and that they are a different means of reasoning entirely. This suggests that my account of literary thought experiments is vulnerable to the no-argument argument, while Carroll's is not. In response I will borrow from both Norton and his opponents, and say simply that thought experiments can be reconstructed as arguments, though they are not themselves arguments. That a thought experiment is not an argument does not make it less useful. That the structural features of thought experiments may be compared to those of arguments is thus defeated if we are not prepared to open up the term `argument` to anything that produces knowledge. That thought experiments produce knowledge is a claim that I have argued for in the first chapter.

In response to the no-evidence argument, which charges literature with a failure to make arguments which can then be debated and analyzed, Carroll suggests that the objection mistakes the purpose of thought experiments. He denies that evidence is relevant to thought experiment because "A philosophical thought experiment is not a device for reaching empirical discoveries but for excavating conceptual refinements and relationships"[158] I do not agree with Carroll that thought experiments are never devices for empirical discoveries, as scientific thought experiments have played important roles in such discoveries. For instance, Einstein's beam of light experiment has never been

158Carroll p.7

carried out. However it was, and continues to be, a reason for scientists to accept special relativity. Carroll likely has philosophical rather than scientific thought experiments in mind when he writes that thought experiments are not empirical evidence. However, I would add that this does not mean that a thought experiment cannot be a device for reaching empirical discoveries. I believe that a small change to Carroll's conclusion, which is that "...it should be clear that philosophical thought experiments, examples, and counterexamples are not vulnerable to the banality argument, the no-evidence argument, and the no-argument argument, since philosophical thought experiments are not aimed at discovering empirical knowledge"[159] A thought experiment does not need to confirm a truth, just illustrate it. To say that thought experiments *are not empirical evidence* rather than 'are not aimed at discovering empirical knowledge', would not impair the strength of Carroll's arguments against the three objections, and would better account for thought experiments in science.

Having dealt with the most common objection to the idea of literary thought experiments, I will introduce a theory based on family resemblance to identify when fiction is a thought experiment. Thus far all that I have attempted to show is that it is possible to have literary thought experiments. In adopting McComb's criteria we have the means to judge when a work is a thought experiment.

Literary Thought Experiments: It's All in the Family

I have argued that literature can be thought-experimental, though not all literature is thought-experimental. Thus we require an evaluative principle to distinguish those

159Carroll p.8

literary works that do function as thought experiments from those that do not. In the introduction I noted that thought experiments are hard to define, and thus far I have focused on function, not definition. However, some way of delineating the concept is desirable as otherwise we risk being pushed toward the unpalatable position that all literature is thought-experimental. In "Thought Experiment, Definition, and Literary Fiction," Geordie McComb uses a cluster concept to identify literature that is thought-experimental. McComb draws on Wittgenstein's account of 'family resemblances' to characterize the "overlapping and crisscrossing similarity relations with each other at different levels of generality"[160] that thought experiments share. For McComb, something is a thought experiment if it has a high enough ratio of these similarities to dissimilarities.[161] Thus a literary object, such as a narrative, is a thought experiment "if and only if that object stands in enough of those relations which comprise the network associated with the cluster concept, and which contribute to the object being what it is, relative to others not in this network." [162] The fewer dissimilarities with this network a narrative possesses, the greater the extent to which it fits in the cluster concept of 'thought experiment'. In sum, a narrative is a thought experiment if it possesses enough of the cluster of 'thought experiment' relations, and will be a thought experiment to a "'greater extent' than another if and only if it instantiates a *higher ratio* of relevant family resemblances to dissimilarities."[163]

In specifying his cluster concept McComb proposes five dimensions of similarity.

---

160 McComb p.208
161 McComb p.209
162 McComb p.209
163 McComb p.209

Thought experiments typically 1) involve a hypothetical, 2) involve an imaginable, 3) involve our own activity, 4) involve a proper cognitive upshot, and 5) involve no empirical justification. [164] The hypothetical (1) is that element of the scenario described in the thought experiment that we entertain without accepting it as true. The imaginable (2) is the element of the narrative that can be imagined using a sensory modality. It might be something we visualize or imagine hearing, tasting, feeling, or smelling. By saying that a thought experiment involves our own activity (3) McComb means that the reader of the thought experiment manipulates or reasons about what is imagined. He explains that "[w]e act more like an experimental physicist manipulating apparatuses and developing theoretical interpretations than an astronomer making observations."[165] The proper cognitive upshot (4) is the application of the scenario to something beyond itself. The scenario is meant to justify or clarify some specific and distinctive claim about the world. That the work in question involves no empirical justification (5) means that the narrative does not employ new empirical data for justification. Again, the thought experimental character of a narrative rests on the degree to which it resembles other thought experiments in sharing these features and lacking other over-riding features that are in some sense antithetical to thought-experimentation.

Interestingly, McComb's theory is similar to mental modeling approaches. There are clear continuities between McComb's account and Miščević's theory of mental model-ling. I argued in Chapter 3 that Miščević's account is very well-suited to scientific thought experiments but is incomplete because, in placing emphasis on visual-spatial

164 McComb p.209-210
165 McComb p.210

manipulation and chains of causation, it does not fully account for ethical thought experiments. Here I will suggest that by reinterpreting and extending the scope of criteria two and three, Gendler's schematic account can be included in McComb's cluster concept.

McComb's idea is very similar to mental modeling approaches, especially Miščević's. According to the third criterion, the involvement of our own activity, McComb intends a quasi-experimental procedure, one we follow by manipulating or reasoning about what is imagined, not just by entertaining it. The active nature of McComb's account of thought experiments can also be seen in his insistence that thought experiments are experienced using a quasi-sensory faculty (criterion 2). This, I believe, shows it to be most closely aligned with Miščević's account, which emphasized the importance of imagining concrete quasi-spatial scenarios—a mental analogue of computer assisted design.

McComb's definition of manipulation better captures the visual-spatial manipulation of mental models of Miščević's theory than it does Gendler's use of schemas to explain ethical thought experiments. Gendler presented schemas as an explanation for how thought experiments can cause people to apprehend ethical issues according to the terms of the thought experiment—to reframe their experiences using the thought experiment as a guide. Ethical thought experiments are not always as visual-spatial as are thought experiments such as Galileo's falling bodies. It is less clear that imagining the people behind Rawls's Veil of Ignorance is fruitful—the ability to imagine how spatial objects relate may be less useful for non-scientific thought experiments.

That McComb does not intend his account to include Gendler style thoughts

experiments can be seen in his treatment of Catherine Elgin's theory of literary thought experiments. Elgin argues that literature can act as thought experiments (and indeed be thought experiments) by allowing the reader to work out the consequences of various scenarios and thus identify features salient to the thought experiment that can be extended to real-world situations. She claims that "[Literary fictions] advance understanding by exemplifying features and playing out their consequences. They constitute imaginative settings in which particular constellations of features are salient and display their significance. They thus afford reason to think that we would do well (or, in some cases, badly) to consider such features salient elsewhere."[166] In response McComb writes, "However, for all Elgin says, such thinking or considering need not involve any *manipulation* or *reasoning* about imagined objects."[167] The question that is raised by McComb's objection is what precisely it means to do an experiment in your mind. Working through consequences and identifying those features and concepts that are important and relevant seems to me to be both reasoning and manipulation. If you were to imagine a hypothetical in which one feature changed, and to work through the consequences of each variant, that would *be* basic experimental method. The salient features act as dependent variables, the scenario as the independent variables, and by running through the consequences of each set of variables in your mind, you imagine the results of the experiment. Simply identifying the salient features is surely an act of reasoning, and working through various consequences while altering certain variables *is* manipulation. By this reading, criteria 2 and 3 could be understood as fitting Gendler's schemas into McComb's cluster concept. Rather than limiting thought-experimental

166 Elgin (2007) p. 47 as quoted in McComb p. 215
167 McComb p. 215 italics in original

procedure to manipulating visualized objects in space, our own activity (3) can be seen as the act of creating mental model and seeing how the salient features produce the cognitive upshot; the use of sensory modalities (2) may be read broadly enough to encompass all forms of imagination. We don't want ethical thought experiments to be relegated to marginal status, and a broad reading of McComb allows for the inclusion of Gendler's schematic account.

McComb attempts to use family resemblance to overcome the problem of extraneous features in fiction. The problem is that in literature there are many elements that do not clearly add to the thought experiment, and may indeed take away from it. For example the love interest in *1984* does not add to the thought experiment about how communication technologies could allow for populations control by a totalitarian state. McComb's theory has the advantage of accounting for elements that are irrelevant, or even antagonistic to a reading of fiction as thought experiment. By treating narratives as having the capacity to be more or less thought-experimental, a narrative that has non-thought-experimental elements may still be read as a thought experiment, though it will be less of a thought experiment. However, the means of identifying family dissimilarities will be problematic. Novels are complex and layered, and different interpretations will lead to disparate identification of elements as family resemblances and family dissimilarities. McComb claims that his cluster concept can provide a framework for discussing and understanding the relation between individual works of fiction and thought experiments in general. Having the capacity to describe narratives as definitively thought experiments or only borderline cases is a strength of this approach, and one that

is especially useful for fiction.[168]

McComb's thought experiment cluster concept may be especially well suited to literature, but should also work outside of fiction, to more generally evaluate the degree to which things are or are not thought experiments. What I want to keep here is the idea that the similarities to thought experiments are as important as the dissimilarities, and that only certain parts of fictional works are embedded thought experiments within larger works comprised of other extraneous elements.

Conclusion

In this chapter I argued for the existence of literary thought experiments. Davenport's account provided a basis for literary thought experiments rooted in the tradition of scientific thought experiments, and Carroll offered an account that emphasized the benefits of literary thought experiments for ethical understanding. Because not all literary works are thought experiments I introduced a family resemblance theory by McComb that provides the means to identify fiction that is thought-experimental. The two points to hold onto from this chapter are the conviction that literature can be thought-experimental, and moreover that the extended narratives that are distinctive of literary thought experiments may be more useful for ethical understanding than

168 McComb suggests a few novels as potential thought experiments. Of these, two, Aldous Huxley's *Brave New World* and George Orwell's *1984,* I would consider science fiction. McComb also discusses David Davies' article on literary thought experiments, in which Davies argues that some fiction is elaborated thought experiment. Davies also picks out *1984* and *Brave New World* as exemplars of how fiction can be thought experiments saying that "writers of utopias or dystopias such as *1984* and *Brave New World* plausibly intend that, as a result of the receiver's making-believe the content of the narrative, she will come to believe that this is how certain societies would turn out, and will therefore amend her views about the merits of alternative political or socio-economic systems." (Davies (2007) p. 33 as quoted in McComb p. 216) I will return to the idea of science fiction as thought experiment, and the influence of utopia in the final chapter.

philosophical thought experiments.

Chapter 7 Science Fiction

Introduction

I considered the use of literary fiction as thought experiments in the preceding chapter on narratives. Having established that some literary fiction can function as thought experiments, it remains to be shown why science fiction is particularly useful for bioethics. Having argued in the preceding chapter that literary thought experiments are possible, the question is 'why science fiction'? To begin to answer this question I adapt a quote from Ursula LeGuin. She wrote that what science fiction does is "extrapolate imaginatively from current trends and events to a near-future".[169] I will broaden this description to say that science fiction uses the imagination to create worlds which explore science and human activities as they could manifest in our future. This is not a definition; rather it is a rough boundary around the kind of science fiction that I want to consider as thought experiments in bioethics. The characteristics of such fiction are that it is about ideas, and is interested in future technologies and sciences, or future consequences of current technologies or sciences.[170]

Science Fiction, Speculative Fiction and Fantasy

This section has two aims: first, to establish a working definition of science fiction for the purposes of this thesis; and second, to show that this definition is compatible with literary thought experiments and includes the element of science to pair with the ethical possibilities of literature in order to establish the aptness of science fiction thought

---

169Ursula K. LeGuin As quoted in Atwood p. 5

170I take the term human technologies in the widest possible sense. Controlling fire is, in a sense, a human technology, and if there were a piece of science fiction that explored Parks Canada's use of controlled burns and suppression of naturally occurring forest fires in National Parks, then this would, for my purposes, be science and technology. Essentially, I mean nothing more than human intervention in the world.

experiments for bioethics. To begin I will define what I mean by science fiction.

The best known Canadian author of literary and science fiction is Margaret Atwood, who has also written about science fiction as a genre. Atwood describes what it might mean for a narrative to count as science fiction in the introduction to her book *In Other Worlds: SF and the Human Imagination*. What is of interest is that she finds no clear means of distinguishing what is from what is not science fiction. About the definition and borders of science fictions Atwood writes,

> Is this a corral with real fences that separate what is clearly "science fiction" from what is not, or is it merely a shelving aid, there to help workers in bookstores place the book in a semi-accurate or at least lucrative way? If you put skin-tight black or silver clothing on a book cover along with some jetlike flames and/ or colourful planets, does that make the work "science fiction"? What about dragons and manticores, or backgrounds that contain volcanoes or atomic clouds, or plants with tentacles, or landscapes reminiscent of Hieronymus Bosch? Does there have to be any actual science in such a book, or is the skin-tight clothing enough? These seemed to me to be open questions.[171]

The questions that may interest us are, first, whether there is indeed a meaningful category of 'science fiction', or a delineation that can be made among works that are science fiction, fantasy, or literary fiction, and second, whether the elements that are typical or stereotypical of such a category are definitive. For the purposes of my thesis, those fictions that contain possible scientific advances will be most relevant, and this is a trope of what we may call science fiction, though it is also found in literary fiction, mystery, harlequin romances, and potentially any other literary genre. For the purpose of thought experiments in bioethics, the examination of possible science as it will affect human bodies and health is most relevant.

171Atwood p. 2

It is plausible that the purpose and form, and not the elements, of a fiction define the category to which it belongs. In the case of the differences among science fiction, speculative fiction and fantasy, the degree to which fiction expresses genuine possibilities may be definitive. These views are expressed in a dialogue between Atwood and science fiction (and fantasy) author Ursula K. LeGuin. In a 2009 book review LeGuin writes, "To my mind, *The Handmaid's Tale, Oryx and Crake* and *The Year of the Flood* all exemplify one of the things science fiction does, which is to extrapolate imaginatively from current trends and events to a near-future that's half prediction, half satire."[172] This is not, of course, a definition of science fiction. It identifies one of the things that science fiction does: namely, examine the world as it is and make up stories about where current trends may take us. This description fits with what mental-modeling theories of thought experiments pick out: the use of the imagination, or of mental faculties, to judge what will happen within certain given parameters, with the key word *imaginatively* standing in for active construction and manipulation.

Atwood responds to LeGuin by making a distinction but, again, not giving a definition of science fiction. In this case what is identified in science fiction is the use of impossibilities-- things that 'could not possibly happen':

> What I mean by "science fiction" is those books that descend from H.G. Wells's *The War of the Worlds*, which threatens of an invasion by tentacled, blood-sucking Martians shot to Earth in metal canisters-- things that could not possibly happen-- whereas, for me, "speculative fiction" means plots that descend from Jules Verne's books about submarines and balloon travel and such-- things that really could happen but just hadn't completely happened when the authors wrote the books... In a public discussion with Ursula LeGuin in the fall of 2010, however, I found that what she means

172As quoted in Atwood, p. 5

> by "science fiction" is speculative fiction about things that really could happen, whereas things that really could not happen she classifies under "fantasy." Thus, for her--as for me--dragons would belong in fantasy, as would, I suppose, the film *Star Wars* and most of the TV series *Star Trek.*[173]

Thus the question of what could happen, in a way like the question of possible and impossible worlds in philosophy, seems to be at the heart of the question for both LeGuin and Atwood. And indeed if we want to use such fiction in thought experiments, then this may be an important distinction. If we can identify the science that is of interest to our bioethical question as being utterly impossible, then the only likely use for it may be to form the basis for a claim of contradiction, as applied in a thought experiment. If, however, positive claims are to be made, genuine possibility of relevant features is called for. This will only be relevant when what is in question, for instance the possibility of human genetic modification, is utterly impossible. If we are responding to a bioethical debate about the societal effects of allowing genetic engineering, a story in which humans who are genetically modified all spontaneously turn into pigs or writing desks will not be relevant, whereas a story in which the effects of genetic engineering seem plausible, though otherwise the story world is populated by fairies and faster-than-light travel and *deus ex machina* (contrived solution) plot twists, may be perfectly reasonable to introduce into the debate.

Perhaps a simple requirement of a good thought experiment is that it is clear whether or not what it proposes is possible, and in what sense it is possible. So to use the *Star Trek* transporter properly in a thought experiment, it falls on the thought experimenter to consider if teleportation is possible, and whether the background science as required by

---

173Atwood p. 6

the parameters of the thought experiment is sufficiently clear. Thus if a transporter is used in a thought experiment only as a method of transport, the mechanics of it may not matter. But if it is a question of personal identity to which the questions of whether you are literally moved, particle by particle, whether a copy is made, whether a copy is destroyed, etc. matter to the subject at hand, the mechanics will matter, and possibly so too will the question of whether this really is something that could happen. I will introduce a theory of analogical models by Mary Hesse later in this chapter, according to which it is not the possibility of the world as a whole that is relevant but only the possibility of those elements that are being transferred to our own.

I think it is fair to say that it will be difficult to precisely define science fiction, with necessary and sufficient conditions. Instead I propose, for the purposes of my thesis, to call fictions that have novel science in them, and deal with actual possibilities, science fiction. This is not an exclusive characterization- if a story concerns itself with the possible applications of nanotechnology, but also includes dragons and ghosts, I will include it for the elements that are useful, rather than exclude it for the extraneous material.

Possibility

The question of possibility, and of far-fetched or 'sci-fi' thought experiments was brought up by Goodenough. His position on impossibility is particularly set against both literary thought experiments and science fiction thought experiments. That some of the situations described in bioethical thought experiments are unlikely may be a problem for these thought experiments. As a field of applied ethics, there is no obvious reason why

bioethics should take notice of impossibilities. An obvious response is that physics is an applied science, and no one thinks that nomological impossibility is a problem for Einstein's beam of light thought experiment.[174]

Does unlikeliness or being far-fetched create special problems for thought experiments? Goodenough thinks the answers are different in science and in ethics. In science, a strength of drastically unlikely thought experiments is that they work even when it is not actually possible to do experiments or to obtain empirical evidence. The famous case of Schrödinger's cat is an example of such an unlikely thought experiment. Galileo's thought experiments using vacuum are a different kind of example, because they are now possible to actually conduct, but at the time the thought experiments were proposed they were technically impossible. A description of Galileo's experiment on the *Semiophysics* website highlights the technological impossibility of this thought experiment: "To the Ancients, a vacuum was unthinkable. But Galileo conducted a thought experiment. Archimedes had shown that what makes lighter objects sink slower or even float while heavier objects sink quicker depends on the density of the medium. Galileo reasoned that if a medium got less and less dense and in fact became a vacuum, then a heavy and a light object falling through a vacuum would fall at the same rate. It was decades after he died that someone was able to create a vacuum and prove Galileo correct."[175] Goodenough considers the kind of impossibility contained in Galileo's thought experiment and approves of it.[176] Goodenough claims that impossible thought experiments are good ones in science, saying that "The scientific use of thought

---

174Brown (1991) p.29
175Leri, webpage http://www.semiophysics.com/SemioPhysics_Articles_mental_3.html
176Goodenough p. 7

experiments is then limited by the empirical nature of science."[177] He bounds the use of impossible thought experiments, claiming that it is always preferable to actually carry out thought experiments as real experiments, and that the qualified nature of hypotheticals drastically limits the applicability of thought experiments. However, Goodenough accepts both Galileo's vacuum thought experiment, which as we have just seen was a practical impossibility at the time it was created, and which makes positive claims. Goodenough does not account for the positive claims made by Galileo. Because this is a destructive as well as a constructive thought experiment, it cannot be only *reductio ad absurdum* that is acceptable. It seems to me that there is reason to think that in science, unlikeliness is no strong charge against a thought experiment.

Goodenough especially deplores impossibility that is to be found in literary thought experiments. As an example of this type of impossibility, Goodenough indicates the smile of the Cheshire Cat's in the book *Alice in Wonderland* by Lewis Carroll. The Cheshire Cat's smile, which is left hanging in the air after he is gone, is an example of an impossibility that is problematic for thought experiments in bioethics: "Carroll's narrative only works for us because our imagining takes place at a superficial level, just sufficient for us to enjoy the story. Once we try to imagine in more detail, the conceptual difficulties eventually emerge. So at the very least, that a hypothetical situation appears to be imaginable is no guarantee that it is in any sense possible."[178] I question the use of this particular example for the simple reason that the textual purpose of the Cheshire Cat is in part to explore ideas of the impossible. In *Through the Looking Glass*, the Queen

177Goodenough, p. 7
178Goodenough p. 9

says, “Sometimes I've believed as many as six impossible things before breakfast.[179]” When the Cheshire Cat appears to the Queen of Hearts and is sentenced to death, the Cheshire Cat causes its head to appear without its body, which leads to a drawn-out argument about whether it is possible to behead a disembodied head. As the text is explicitly using the Cheshire Cat to explore questions of impossibility, the conceptual difficulties are not an issue of superficial plausibility that quickly gives way once serious attention is applied.

Of course, that this one example of impossibility may be discounted because the impossibility is part of the original intent of the work does not mean that other narratives will not contain problematic impossibilities. For instance, a good number of impossibilities can be found in pulp time travel stories. And so, objections aside, Goodenough's point is well taken as there do seem to be fictional accounts that are imaginable on first inspection, but which we cannot compass once we try to imagine them in detail. That the imaginable and conceivable are not coextensive is not a claim that I intend to contest.

This kind of impossibility needs a resolution, and is connected I think to Goodenough's point about under-description in narrative.

> A thought experiment, of course, is not a situation but a description of a situation, a narrative, and that carries with it its own possible problems. There is, for instance, the problem of under-description. Now in ordinary science fiction the audience doesn't usually care about the fine details of some hypothetical device. For the purposes of narrative enjoyment, we swallow Star Trek's transporter room, even when an episode's plot raises some apparently philosophical problem raised by

179"Definition of cheshire cat." *Webster's Online Dictionary*.

> its malfunction. (Two Captain Kirks? Two people or two tokens of the same person?) But when a philosopher like Parfit asks us to take such an example seriously, it should be hard for us to come to any kind of intuition about the output of such a device until we have some reasonably clear idea of what it actually does.[180]

First it is worth noting that real-life situations as they are presented in case studies are also narratives which will contain only those elements that the author considers relevant, and which are generally not available to later readers' examination beyond what is presented in the text. Thus, real cases are not immune to this criticism. There are certainly special problems in science fiction, as Goodenough points out using the Star Trek transporter. The mechanics of teleportation may matter for identity theory, and are never fully explained in Star Trek. However, this charge would apply equally to Parfit's thought experiment in which people split into two individuals identical in physical form and consciousness. That there are potentially problematic thought experiments not drawn from science fiction is, of course, no defence of science fiction thought experiments. And indeed, whether or not such thought experiments are a problem is by no means certain, and at the end of this section I give a brief defence of such thought experiments.

A lack of relevant information can affect case studies as well. For example, insulin coma therapy was widely used in the 1940s and 1950s for psychiatric illnesses including schizophrenia. In 1953, Harold Bourne published a paper titled "The insulin myth," which denied that there was any evidence that using insulin shock to induce comas had any therapeutic benefit, and that the purported results were the effect of psychiatrists choosing the patients with good prognoses for the therapy and the effects of better care for these patients. Earlier case studies did not provide the relevant information to show

---

180Goodenough p. 9-10

what the salient features of the treatment in fact were. In any reporting, not all of the facts, and possibly not all of the relevant facts, will be described. Under-description is thus a problem that extends beyond thought experiments.[181]

It is possible that science fiction is particularly vulnerable when we rely on science fiction thought experiments for mechanisms that are not available. Goodenough's claim that an "audience doesn't usually care about the fine details of some hypothetical device"[182] may be true for a majority of the audience, though it fails to account for a devoted geekdom that is passionately interested in the fine details of hypothetical devices of science fiction.[183] It is worth noting that there is a substantial audience interest in the possible workings of science fiction devices. Authors of hard science fiction have in some cases made robust attempts to explain devices such as spaceships in scientifically plausible ways. One example is the science fiction writer Catherine Asaro, who holds a PhD in chemical physics and writes novels set in a future that includes faster-than-light travel. Asaro describes her ideas on the Physics Central website: "You can't have a galactic empire without a way to go faster than the speed of light," says Asaro. "So I wanted to come up with a believable way to do it, even if it's not physically possible." The solution Asaro found was a mathematical trick involving imaginary numbers. "It's as if you're traveling in the complex plane," she explains. "It's actually kind of simple, and

---

181 Please also see the section on narrative and mental modeling in Chapter Two for some thoughts on under-description in literature.

182 Goodenough p. 9-10

183 A Google search for the terms "star trek transporter how it works fan" produces 13,300,000 results, the first page of which includes two detailed wikipedia sites, a fan created Frequently Asked Questions about Star Trek Tech, an essay posted on the site Common Sci-Fi Debating Tactics and Fallacies and three other Star Trek fan sites.

it's pretty."[184] Asaro published these ideas in much greater detail in the article "Complex speeds and special relativity" in the *American Journal of Physics*. The Star Trek transporter is under-described, and any philosophical use of this device that relies on an understanding of the mechanics of teleportation does face serious problems. However, not all science fiction devices are under-described, and it is not clear why a thought experiment from a well-described device such as Asaro's means of travelling faster than light would face more problems than any thought experiment that used ideas from theoretical physics more generally.

It may also be useful to distinguish when it is appropriate to object to impossibility. In the movie *Waterworld* global warming causes the sea levels to rise so high that the only dry land is the tip of Mount Everest. This is impossible because the complete melting of the ice caps and glaciers would not produce enough additional liquid water to raise ocean levels by approximately 150 meters[185]. This kind of impossibility seems to me useless because it fails to capture any concept of interest. The movie *Tank Girl* is also set in a world of extreme climate change, but with the premise that water became a sufficiently valuable commodity that corporations have managed to contain and commodify all of the water on earth. This seems equally impossible given the current state of the world, but is a conceptually interesting impossibility. It is not currently possible for any human organization to contain all of Earth's water in such a way, but questions about human rights and access to water make it an interesting thought experiment in a way that the

184http://www.physicscentral.org/explore/people/asaro.cfm

185"The Real Waterworld." *HistoricalAtlas.com: the Centennia Historical Atlas -- Europe and the Middle East* . N.p., n.d. Web. 23 May 2012. <http://www.clockwk.com/waterworld/index.html>

miraculous multiplication of water on earth is not.[186]

This idea is similar to a response Simon Beck makes to problems of impossibility in personal identity, in which he denies that irrelevant impossibilities detract from thought experiments. In "Should We Tolerate People Who Split?" Simon Beck considers the degree to which unlikely, or I might say 'sci-fi' thought experiments, are made irrelevant if they are currently impossible; not logically impossible, but factually impossible. What is important is to identify whether the impossibility in question in fact affects the question at hand. Beck writes that "If our current, present notions allow splitting, albeit alien, persons then it is not at all clear that such things are irrelevant impossibilities. The impossibility, if it is one, of human fission would rather seem to be the impossibility which is irrelevant to the outcome of experiments in which the aim *is to find out about persons."*[187] Beck is writing in response to an argument by Wilkes which identifies three problems with impossibility in thought experiments: that we lack relevant background information, that the ability to imagine something does not make it possible, and that unlikely thought experiments are simply too distanced from reality to be of use.[188]

In response to the problem that 'we are not given the relevant information', I have in this section argued that case studies are not immune from this problem. In addition, in the previous section on narratives I made the point that thought experiments taken from literature are actually richer in background detail than other thought experiments, and that this allows for more finely grained ethical understandings. That 'what one person

---

186It could be said that given the potential of global warming to sink coastal cities that a thought experiment about radical landscape change due to climate change is a useful thought experiment.
187Beck p. 11-12
188Brown p.29

finds intuitively certain another will consider obviously false' is related to the status of moral intuitions, and which I concluded only creates problems for certain thought experiments, namely those with a justificatory purpose. The status of moral intuitions is not a problem when these intuitions are used as information to parse concepts, or as motivation for actions which we can argue the merits of without reference to the intuitions elicited using thought experiment. As for the charge that 'thought experiments take us too far from the actual world' and that 'the fact that we can 'imagine' something doesn't mean it's possible', I believe both come back to the question of how much context is available. Brown responds that "Thought experiments often involve a kind of counter-factual reasoning, yet counter-factual reasoning is extremely sensitive to context."[189] In addition, as I discussed in the previous section on narrative, literary thought experiments are often richer in context than case studies simply because a novel provides more scope for detail than does a typical case study. I will next give some reasons for adopting science fiction thought experiments in bioethics.

Analogues and Analogies

The problem with possibility for science fiction is creating an account that separates relevant from irrelevant impossibility and from distant possibilities. An account is required which picks out the relevant similarities when there are many irrelevant differences. For example, the books of Lois McMaster Bujold are set on other planets which are in many ways dissimilar to Earth. One of the ideas in her novels is the effect that artificial wombs would have when introduced to a historically patriarchal society. The salient features for understanding this as a thought experiment which models the

189Brown (1991) p.30-1

societal changes related to freeing women from childbearing are not that the planet is supposed to be located far from Earth, that the colonization of the planet is an artifact of a history that we do not share, that the flora and fauna of the planet is different than ours, and so on. The total similarity or dissimilarity of worlds is not what is at issue—the questions of possibility and relevant similarity should encompass the social structure of the proposed society, the social changes resultant from artificial wombs, and the consequences which are portrayed for characters insomuch as they reflect the effects of the novel technology. It is not the possibility or likelihood of the fictional world as a whole that is relevant, but only the possibility or likelihood of those elements which are salient to the thought experiment.

Analogical arguments offer a means of understanding how possibility can matter for only some aspects of the fictional world. Analogical arguments conclude that a target object has a certain property *x* on the basis of its sharing a set of related properties with a model object that has *x.* Not all of the properties of the target and model must be shared. In the case of applying literary thought experiments to real world situations, the possibility of the world as a whole is irrelevant. Only those aspects of the world which are imported to our own by the thought experiment—those which are implicated in the cognitive upshot of the thought experiment or, in other words, are used in building our mental model—must be relevantly similar to the real world. Analogical arguments do not require that all aspects of the compared objects match. Indeed, if they did, it would be identity and not analogy. Identifying which common properties are implicated in the analogy can be likened to identifying the elements of a fictional narrative that are implicated in a thought experiment. In *Models and Analogies in Science*, Mary Hesse

distinguishes among positive, negative and neutral analogues. Positive analogues are properties that are shared, negative analogues are defined as properties that are not shared, and neutral analogues are properties for which it is currently unknown whether or not they are shared. When an analogue is essential to an analogical argument, but it turns out not to be shared by the model and explicandum then the analogical argument is refuted.[190] Hesse denies that there is a set criteria for judging when properties are essential, but, roughly, essential properties are those properties causally or conceptually related to the property of the target object that the argument is meant to establish. Hesse maintains that the fundamental requirement of a successful analogy is that "The essential properties and causal relations of the model have not been shown to be part of the negative analogy between model and explicandum."[191] This can be applied to thought experiments by treating the elements of the fictional world as analogues, divergences between the fictional and real world as negative analogies, similarities as positive analogies, and the thought experiment results as neutral analogies. Thus Bujold's planetary history would be a negative analogy, the social structure of historical patriarchy overlain with new equality would be a positive analogy, and the imagined results of artificial wombs would be a neutral analogy. And, as in Hesse's theory, it is only when the negative analogue is essential and causes the neutral analogue to become a negative analogue that it is fatal to the analogy. So only when the differences between the fictional and real world make the thought experiment inapplicable does the fictional thought

190Hesse p. 90
191Hesse p. 91

experiment fail because of those differences.[192]

The point to draw from Hesse is that it is only those ideas implicated in the cognitive upshot of a thought experiment that must be shared between the fictional and real worlds. Thus for thought experiments drawn from science fiction works which include bizarre, impossible, unlikely or unimaginable aspects, if the weirdness is unimportant for the content of the thought experiment, then the impossibility is not a problem for the thought experiment. When the weirdness is implicated in the thought experiment itself, and especially when the weirdness affects the cognitive upshot of the thought experiment, or affects the model created by the thought experiment, then Hesse's rules for essential properties give us a way of evaluating the degree of trouble the weirdness is causing in the thought experiment.

Utopia

In the previous section on narrative, all of the accounts defending literary thought experiments included Orwell's *1984,* and also referred to are Huxley's *Brave New World,* Thomas More's *Utopia,*and Wells's *The Sleeper Awakes.* These novels were included as paradigmatic thought experiments. My reason for discussing utopias is threefold: because they appear frequently in the discussion on literary thought experiments, because they

192The problem with using Hesse's theory of analogy is that while she believes that analogies are used in creating models that advance novel predictions in science, she treats analogies as arguments. Analogies are distinct from thought experiments, and I have argued that thought experiments are not arguments. Because analogies are not thought experiments I will not re-argue the case that thought experiments are not arguments. Instead I will say simply that analogies as models provides a useful starting point for separating relevantly close possibilities. However, I do face one large problem—Hesse treats analogies as methods of selecting hypotheses which dispenses of the problem of justifying analogical models as invalid inferences. In the first chapter, I argued that thought experiments produce knowledge, and do more than offer means of coming up with creative hypotheses. Thus it seems that I have the problem of explaining why thought experiments are not invalid deductions. However, this is only a problem if thought experiments are arguments, and if I am right, they are not arguments. What I agree with Hesse about is models, and what I want to take from her theory is the means of selecting relevant shared features.

share characteristics and overlap with science fiction, and because the particularly moral characteristic of utopias is useful in justifying why science fiction thought experiments are apt in bioethics. It is this third reason, the moral characteristic of utopias, that is most important for my thesis because showing that some (utopian) science fiction has moral elements will help me argue that science fiction is useful as bioethical thought experiments.

As a genre utopias and dystopias propose social structures that diverge from what is in actual practice. Ideologies, including communism, anarchism and libertarianism may be used in utopias. However, the way that they are depicted and interact with other elements of imagined societies is unique. Ethics is built into utopias, as it is built into philosophical debates on political theory. The ideas of fairness, justice and the good are of course normative in character, and are implicated in utopias simply by the nature of the task, which is to imagine the best of all possible societies, or, by considering the worst of possible worlds to identify what it is that makes it so. If it can be argued that utopias are related to science fiction, then this moral character can be said to underlie science fiction as well as utopias.

What is characteristic of utopias is the examination of what could be different about society. This is of course the central motif and identifying feature of both utopias and dytopias and falls very much into the tradition of seeing literary thought experiments as narrative enactments of a hypothetical. Utopias are often described as didactic, which is in itself an interesting term, and one lighted upon in the earlier discussion of literary thought experiments. It is an odd thing that we view instructional purpose in literature as

something singular and separate from the general purpose of art. Atwood writes that "The curious thing about serious utopias, as opposed to the satirical or entertainment variety, is that their authors never seem to write more than one of them; perhaps because they are products, finally, of the moral rather than the literary sense."[193] And this does seem to be reason to view them as thought experiments. Indeed, by their character as examinations of what could be different about society, utopias seem to be detailed and extended thought experiments in political philosophy, which is an idea explored by Miščević.

In "Political Thought Experiments from Plato to Rawls," Nenad Miščević argues that utopias and dystopias are thought experiments. The reason for accepting utopias as thought experiments is that the literary and philosophical utopia have the same origin and are structurally similar. The process of running a thought experiment and reading a utopia is described in the following passage as very much the same; both involve engaging the reader to imagine the described world, and to use their background knowledge and ethical attitudes or intuitions as well as the possibilities they envision to draw conclusion about the proposed world, and so about the social set up of that world.

> Interestingly, one could reconstruct the standard experience of the reader of political fiction, the literary utopias (including dystopias), in roughly the same stages, but involving some literary devices ... And the readers go through the same process of asking themselves, probing their "moral sense", and in the case of negative utopia coming up with a critical intuition: "No, this arrangement is morally disgusting". ... Issues, examples, reactions fall into a pattern (the Big Brother turns out not to be a brother at all, the ideology of the "beneficial" state starts showing cracks and contradictions, the hero's superiors start showing their nasty character). The components of the big collective arrangement in the story support each other and the negative judgements about most of them counterbalance the few positive instances. One hopes that personal experience and the factual

193Atwood p. 105

> knowledge of the reader join in, leading her to a wider reflective equilibrium.[194]

Miščević connects the tradition of armchair speculation or of ideal theories in political thought experiments with literary utopias. The structure that Miščević proposes is similar to his general theory of mental modeling with a greater emphasis, I believe, on causation and the modeling of effects, and less on visual manipulation of mental models. By separating political thought experiments into two traditions, Platonic ideal states and social contract theories, Miščević connects literary utopias to the former and dystopias to the later. I have some reservations about this approach, as not all dystopias are chiefly interested in what members of society would be willing to agree to[195] and the results of such agreements. This minor point aside, the feature that interests me is that the genre of utopias and dystopias is connected to the tradition of thought experiments not only by their history, but by shared features. The argument is not just that we can use literary utopias as thought experiments, or that the traditions of political thought experiments and utopias are connected, but that what utopias do is what thought experiments do-- that literary utopias are thought experiments. Miščević identifies this as the stronger version of his arguments, saying that "The more ambitious hypothesis would connect the two clearly philosophical TE-traditions, the Platonic and the social contract approaches, with their positive utopian and negative-utopian outgrowths, from More, Campanella, and Harrington to authors like Fourier and Owen all the way to clearly literary fictional utopias, for instance the negative ones of Orwell and Zamyatin and the "ambiguous" one

---

194Miščević (forthcoming) p. 197-8

195One example would be *Brave New World.* We would not consent to such a society, however, the inhabitants of *Brave New World* clearly would.

of Ursula Le Guin."[196]

That the structure of utopias is analogous to that of philosophical thought experiments and that they both serve the same function of imagining ideal societies are compelling reasons for reading literary utopias as thought experiments. Even beyond this point Miščević hints at the possibility of creating a deeper understanding of thought experiments through the study of literary forms that have traditions of thought experimentation. Unfortunately Miščević does not elaborate on the point, but he does claim that thought experiments are important in literature as well as in philosophy, writing that "Since TEs are central for other areas of philosophy and for some traditions in fictional writing, once we put the big picture together, we'll get a more unified view of the whole, which could help us to integrate the methodology of political philosophy with methodology of philosophy in general, and finally contribute to a deeper understanding of fictional-literary thought experimenting."[197] This is a claim that I am interested in, especially as it regards science fiction as a genre in which questions about possibilities are explored through the creation of possible worlds, which I would argue makes science fiction one of those traditions of fictional writing for which thought experiments are central.

Miščević makes a good case for utopias being thought experiments. Given the overlap between utopias and science fiction and the way in which both are defined by imagining ways the world could be substantially different than it currently is, that utopias are moral in nature *and* thought-experimental means that utopias do two of the three things required

196Miščević (forthcoming) p. 204
197Miščević (forthcoming) p. 205

for relating science fiction thought experiments to bioethics—they are literary thought experiments and are ethical in nature. All that is missing is the element of human science, and if my definition of science fiction is workable for this purpose we may simply include those works that include science that affects human bodies[198] and exclude the rest. As a type of fiction that is supposed to justify certain beliefs and motivate certain actions about possible futures or social structures utopias serve a cognitive function similar to thought experiments. By providing mental models through narratives they work in the same way as thought experiments. Those which have possible future science or technology shaping the fictional society both bring in aspects of scientific thought experiments and at the same time are clearly science fiction.[199]

Science Fiction Thought Experiments in Bioethics

Science fiction ties together ethics and science just as bioethics does. Science is definitionally included, and ethics, the questions of right and wrong, and ought and who did whom wrong are woven into the fabric of fiction and of narrative. Ethics is implicit in fiction, and science fiction combines science and ethics, just as bioethics brings together science and ethics within philosophy.

If we can agree to a definition of science fiction by which science fiction has to be about science and have moral elements it is clear why science fiction thought experiments are a good fit for bioethics: they bring together science and ethics. In

198This is not to say that all bioethical issues will be about human bodies. For instance Margaret Atwood's novel Oryx and Crake features chickens made in vats: "That's the head in the middle...There's a mouth opening at the top, they dump the nutrients in there. No eyes or beak or anything, they don't need those." (Atwood (2003) p. 202) With technology now promising to allow chicken proteins to be printed onto an artificial scaffold, this and other science fiction that feature 'vat meat' may be forecasts of societal reactions to a very different form of protein production.

199Thanks to my supervisor, Letitia Meynell, for raising this issue.

addition, science fiction in general has narrative characteristics that enhance ethical understanding and engage cognitive capacities associated with imagination and visualization. This imaginative ability is exemplified in the use of science fiction thought experiments such as those that work through a set of plausible possibilities for a technology, and potential moral consequences of its adoption, of a particular use, or of the banning of such a technology.

I am not claiming that science fiction thought experiments are always well used in bioethics. They may be empty, inaccurate, irrelevant or unlikely. Science fiction is often referenced in the wrong way in bioethics. It is clearly useless to say 'look at *Brave New World*. If we allow cloning, the result will be an oppressive society' as an empirical claim about what will happen. But if we say 'look at *Brave New World,* as a counter-example for the claim that technological advancement necessarily promotes individual freedom', then we have a thought experiment that is possible, internally consistent, and potentially useful in bioethics.

Regarding the special applicability of science fiction, I will look at a few works of science fiction that are concerned with a what-if question current in bioethical debates. My argument is that the elements of science and of 'what-if' are what specially suit science fiction thought experiments to use in bioethics. This point is made in the following quote by Julie Czerneda, a science fiction author: “Science fiction is read not only for enjoyment, but because it digs into scientific concepts with imagination, creativity, and a thorough appreciation of consequence. Most science fiction authors ask, "What if" and speculate about what could happen if a certain aspect of science or

technology existed."[200] The area in which science fiction is arguably the most useful as bioethical thought experiments is in novel technologies. Science fiction, by its nature, often deals in science that is not yet in practice. This is to say that it takes as its subject scientific advances that do not yet exist. In many cases of course this science is fanciful, but in some cases the science is now emerging in a form similar enough to that in the fiction to be analogous for conceptual if not scientific purposes.

Examples

One trope of science fiction is human genetic engineering. This topic has also seen a great deal of attention in bioethics. Geneticist Lee Silver works in bioethics, and the epilogue of his book *Remaking Eden,*[201] is written in the form of a science fiction short story. This story describes a possible history for human engineering spanning the coming millennium. Silver warns that genetic modification can lead to changes so great as to be equivalent to speciation not only between modified and unmodified humans, but also between humans modified in different ways. In Silver's thought experiment he makes it explicit that his format is a way of extrapolating the possible and probable results of current science. Whether or not we find Silver's thought experiment plausible, what is interesting is that it was written as a narrative, and one that follows in the footsteps of science fiction. Silver couched his extrapolations as not only a thought experiment, but as a science fiction thought experiment.

The idea of human genetic modification that has drastically altered humans and human society has been taken up in countless pieces of science fiction. I will look at only

---

200 Czerneda p. 39

201 Lee M. Silver, *Remaking Eden: cloning and beyond in a brave new world,* (New York: Avon Books, 1997).

a few here, with the purpose of showing that they have the purpose and effect of imaginatively extrapolating the ethical result of novel human sciences. *Brave New World* by Aldous Huxley is a novel that is ubiquitous in bioethics. I include it here because it is so often cited in relation to genetic engineering and cloning, though the novel includes neither. *Brave New World is* a work of science fiction and also a dystopia, and thus combines moral considerations and human technologies. The most notable technology is artificial reproduction. This novel is a dystopia in which a totalitarian World State has complete control over the populace and over human reproduction, which is done in assembly line fashion, with the exception of small reservations that are not controlled by the World State. State control of reproduction, education, employment and information is complete, and this control is linked to the ability of the state to grow populations to order using artificial reproductive techniques. That *Brave New World* is a thought experiment was argued for in the preceding section on utopias, and also the preceding chapter on literary thought experiments. That this novel has ethical content is implicit in the judgement we make that *Brave New World* is a non-ideal society, and the reproductive technologies that make complete state control possible are an imagined use of a possible development in biological technology.

A less familiar science fiction future is created by David Marusek[202] in the novels

---

202Marusk's short story "The Wedding Album" is a particularly lovely thought experiment about artificial intelligence and personal identity. The premise is that instead of photos, neural patterns are captured whenever anyone wants to capture a memory, and these neural patterns have computational powers that include remembering, describing, arguing, imagining conterfactuals, and making inferences. The story story is about one captured neural pattern from a wedding album, and the experience it has of being evaluated for artificial intelligence after a revolution in which all sufficiently advanced intelligences are declared 'persons' with the rights thereof.

Counting Heads[203] and *Mind Over Ship*[204]. These stories are set in a world that has effective anti-aging technologies, which lead to the passing of laws preventing new births as a means of reducing over-population. The effect is a world full of the rich, creating a market for clones (a procedure which is not covered under the birth prohibition). The clones are made from a small number of individuals, creating 'types', each of whom has many genetically identical instantiations. Thus 'Franks', clones made from the cells of a person named Frank, are manual labourers, security guards and firefighters, 'Lolas' are maids and entertainers, and so on. This is a bioethical thought experiment in that it takes a possible future technology, perfected human cloning done on a large scale in conjunction with life-extension technologies, and models an unexpected outcome based on current conditions and the possibilities of the proposed technologies. This science fiction world also partakes of dystopic elements, and invites a moral condemnation of the future society. This is reflected in a review by Dave Itzkoff for *The New York Times.* Itzkoff writes of the book *Counting Heads* that it is

> an ambitious, sometimes brilliant and sometimes overwhelming attempt to provide a fully realized portrait of what society might be like in the 22nd century, when rapid advances in every field from cloning to artificial intelligence to nanotechnology have made our planet both a simpler and a more sinister one on which to live....At its best, the novel makes a reader nostalgic (if that's the right word) for the present time, and grateful that he will never see a future in which the human body has become devalued to the point where it is merely a storehouse for information, and no one ever really dies — they are coldly declared "irretrievable." As one character sardonically puts it, "No time and no bandwidth — that's about as good a definition of death as I can imagine."[205]

203Marusek, David. *Counting heads*. New York: Tor, 2005. Print.
204Marusek, David. *Mind over ship*. New York: Tor, 2009. Print.
205Itzkoff, n.p.

Again we have a science fiction thought experiment that takes novel bioethics and pictures alternative possibilities, using imagination to visualize and make concrete what could come of the development and adoption of cloning and life extension.

The stories that I have presented so far have all acquired distinctly moral thought-experimental characteristics from sharing in the tradition of utopias. The short story "Conditional Love" by Felicity Shoulders[206] derives its ethical elements from a structure that places the potential harms of imperfect genetic modification in the context of the lives of several vividly described children. In this story regulations exist to prevent untested genetic engineering, but underground providers offer genetic modifications. The story is set in a hospital ward for children who have failed genetic alterations and have been abandoned by their parents as a result. The moral element is found in the sense of injustice the reader is invited to share in individual cases, and not in larger societal structures. The world is no more than sketched, with the focus resting on a few individuals for the entire story. Again, we have a moral element, new biotechnologies, and a narrative that is primarily about the extrapolation of what could have happened: a thought experiment.

A different vision of human speciation resulting from genetic engineering is presented in Nancy Kress' *Beggars* trilogy[207]. Genetic engineering leads to a world where the United States is split into three completely separate classes that do not live or work together. The 'sleepless' are a group of people with a genetic modification that eliminates their need for sleep, and as an unintended consequence also makes them magnitudes

206 Felicity Shoulders, "Conditional Love," Asimov's Science Fiction.

207 Kress, Nancy. *Beggars in Spain a novel*. New York: Avon Books, 1993. Print., Kress, Nancy. *Beggars & choosers*. New York: TOR, 1994. Print., Kress, Nancy. *Beggars ride*. New York: Tor, 1996. Print.

more intelligent and unaging. These traits are heritable, and the society that is formed by the sleepless is almost entirely separate from the rest of humanity. The 'donkeys' are the genetically engineered children of the elite, who have enhanced intelligence and other desired traits. This class also forms enclaves, but has business and other dealings with the non-genetically modified class. The 'aristos' are the vast majority of the population; they are not genetically modified, and are supported by a welfare state. The gulf between the donkeys and aristos is presented as a nearly insurmountable schism because of the differences between them, and the gulf between the donkeys and the sleepless is even greater and more rigidly maintained, both by societal factors and by the differences in individuals' abilities. The idea that is explored in these novels is how the current societal structure of the United States, or of the Western world, would be affected by drastic changes to human capacities brought about through genetic engineering. These novels are not pure dystopias. However, they share a great deal with dystopic traditions in that the main character of the trilogy is the society itself, and the overall picture is of a troubling society in which the rich and poor are increasingly separate and mobility between classes is reduced to non-existence. This is an invitation and framing for moral judgement, and extrapolation of where the divide in haves and have-nots in American society would lead with reinforcement from new technologies that would stack the deck further in favour of the rich in a society that combines meritocracy and contemporary patterns of inheritance of wealth. Thus it brings together ethical judgement, with rich background and context and in developed characters from each of the classes, and novel human sciences.

Having given some examples of science fiction that have bioethical themes, I will now

focus on Kress's novels to show how such narratives work as thought experiments. The first step is to show that they are thought experiments, and only from there can I argue that they are thought experiments that have the potential to be useful in bioethical discussions. McComb's means of distinguishing narratives which are thought experiments requires the possession of 1) a hypothetical, 2) an imaginable, 3) our own activity, 4) a proper cognitive upshot, and 5) that the work in question involves no empirical justification. McComb uses *Brave New World* as an example of a novel that works as a thought experiments. I will now show that Kress's novel also meets McComb's standards for literary thought experiments.

There is a number of hypotheticals in the *Sleepless* novels, including the central idea of genetic engineering that creates people who do not sleep or age. Other hypotheticals relate to the way that the economic, political and class structure of the United States could change. It may be argued that the genetic engineering proposed by Kress is unlikely, but there is no problem imagining people who neither sleep nor age, or an extremely stratified society with a robust welfare state. That the work is science fiction makes it especially easy to establish that there is no reliance on empirical justification—the story is not about the world as it is, but as it may be and so is both a hypothetical, and without robust empirical justification[208]. It is the concepts of our own activity and of a cognitive upshot that are the meat of the issue.

That there is a cognitive upshot to the Sleepless novels is easy to argue, but that this cognitive upshot extends beyond the boundaries of the narrative is a question with greater

---

208 The concept of an imaginable is tied to the section on possibility, and I will not examine it again except to say that being imaginable is a significantly lower bar that that of being possible.

weight. There are ideas about human modification and the potential to exacerbate current oppressive patterns wherein parents who have higher social positions are able to provide environments, education, health care and other goods that make it more likely that their children will hold high social positions. If parents can pay to ensure that their children will also have greater intelligence, physical attractiveness or other social goods, it seems likely that individuals from marginalized and disadvantaged populations will face an additional oppressive mechanism. The question of whether this cognitive upshot has traction outside the narrative is key to my argument. There is a current debate in bioethics about whether or not genetic engineering should be allowed, and I would argue that issues of systemic injustice are relevant to this debate. Thus the cognitive upshot of the *Sleepless* novels has an extension beyond the narratives, into the debate on human genetic modification.

I will look to mental modeling theories for reason to believe that this narrative evokes the readers own activity. The features of unfolding in time and following a specific causal sequence[209] come from Nersessian, and as I discussed in the chapter on mental modeling the features of taking place over time and somehow involving causation are characteristic of almost all narratives. The *Sleepless* novels make use of two devices relating to unfolding in time. The first is following characters through time. The second is establishing a narrative of the society separate from the characters. The three novels each follow different characters, and in each novel the characters are described over a significant, described length of time. This narrative technique causes us to build in our minds a sequence of what happens not only to the characters, but to the society they live

209Nersessian pp.295

in. Following a specific causal sequence is related to the way that the narrative unfolds over time, and what is of most interest is the ways that genetic engineering seems to affect and be affected by the society in which it occurs. By seeing the story play out in time, the causal factors are clarified, and the causal linkages make sense of the story that unfolds. By following the narrative in time and in causation, a mental model is developed, which the reader can then manipulate.[210] Thus we create a model of what happens. Creating and manipulating this mental model is our own activity. Thus I argue that this science fiction narrative has all of the family resemblances that McComb calls for to identify literary thought experiments.

Having established that the *Sleepless* novels are thought experiments, I argue that they are useful in the field of speculative bioethics. To do this I will return to the idea of a cognitive upshot beyond the narrative. There is lively speculation in bioethics over what the ethical and practical ramifications of human genetic modification could be, and based on these theories there have been calls to allow, restrict, or ban human genetic modification. If this thought experiment can create a model for how genetic engineering might lead to increased structural injustice, then this thought experiment plays an explanatory or mediative role in such debates, and thus provides a potential model for how human genetic modification could play out. The four science fiction pieces that I have described here all contain ethical elements, novel technologies, and imaginative modeling of future worlds.

---

[210]Due to the characteristics of narrative, which I discussed in relation to mental models, it is hard to imagine a developed narrative that would not satisfy this criterion.

Conclusion

Thought experiments work as ways of picturing alternate possibilities and of using imagination to grasp actions and their probable outcomes. Science fiction thought experiments are ideal for issues in bioethics because as ethical narratives they are context-rich and concrete, and the speculative nature of science fiction lends itself to the exploration of new and emerging human sciences. By invoking alternative worlds, and through the vivid imagining of what *could be,* science fiction thought experiments fit with mental-modeling theories of thought experiments. Indeed, mental-modeling accounts are exemplified in science fiction thought experiments as they work to exercise the imaginative capabilities and use imagination to grasp possibilities. The ability of mental-modeling theories to encompass works of science fiction gives us reason to think that science fiction in particular acts as literary thought experiments. Thus science fiction is thought-experimental and concerned with bioethical sciences. Utopias serve a similar cognitive function to thought experiments, to justify beliefs and motivate actions. Moreover by using narratives they enable the creation of mental models, and those utopias which engage with questions of possible future science and the shaping of societies are both science fiction and thought-experimental.

Science fiction is thus a good resource for bioethics because it brings together science and ethics, which was manifest in a variety of examples of science fiction through experiments that are concerned with what-if question current in bioethical debates that were presented in this chapter.

Chapter 8 Conclusion

In this thesis I argued that science fiction is particularly apt as bioethical thought experiment. I began by making a case for the use of thought experiments in bioethics, and then argued that fictional narratives can be thought experiments, and that science fiction is a class of fictional narratives that has special applicability in bioethics.

Scientific thought experiments produce knowledge, as I showed by considering two thought experiments by Galileo. These thought experiments share significant features with real experiments, notably that they advance scientific knowledge. In considering the theories of Brown, Norton and Buzzoni, I suggest that mental-modeling theories of thought experiments afford the best explanation because they explain how thought experiments can be like real experiments without literally *being* real experiments, and without the pitfalls of understanding thought experiments as either arguments or as relying on *a priori* truth.

I proposed a version of mental modeling that has the flexible modalities of experience found in Nersessian's account, combined with Miščević's compelling vision of how existing knowledge is used to create mental models, and Gendler's use of schemas to understand ethical thought experiments. Key is the use of imagination to create and manipulate models or schemas, which make use of the breadth of available experience and background knowledge, which is experienced using a range of sensory modalities, and which is explicitly narrative. In emphasizing the importance of narrative in thought experiments, mental modeling gives reason to think literature has a place as thought experiments.

In considering ethical thought experiments I argued that ethical thought experiments unearth our intuitions and explain or motivate ethical actions and that mental modeling, particularly Gendler's account of schematic mental models, makes sense of this. I looked at Thomson's violinist as an explanatory thought experiment, Singer's drowning child as a motivational thought experiment, and Rawls' Veil of Ignorance as a justificatory thought experiment. Ethical 'intuition pumps' are problematized by the status of moral intuitions. However, these problems largely stem from a concern with the status of moral intuitions that is not confined to thought experiments, but is a general metaphysical problem for ethics. Moreover it is only for justificatory thought experiments that moral intuitions are a significant problem, and not explanatory and motivational ethical thought experiments. Bias and variation in intuitions elicited among individuals and among variants on thought experiments are potential problems for ethical thought experiments. Moreover, bias is a problem that mental modeling both points out in the use of chunked information, and fails to give ready remedy for because the use of background information is necessary for mental models. However, thought experiments hold out the promise of clarifying moral intuitions and by using a method like Hintikka's we see that the variability of intuitions is not necessarily a problem, but can be a source of information.

Bioethics makes use of thought experiments' capacity to move from abstraction to discrete instances, which is one of the useful capacities of thought experiments. Moreover, if we give up thought experiments, then we lose elements of ethics and science that are valuable to bioethics. I argued that thought experiments can identify and clarify in ways that case studies may not be able to, and that thought experiments enable

the use of forms of reasoning based on the use of mental models that are not available via argumentation. Sometimes thought experiments will be better, and sometimes real cases will be unavailable. Given the cognitive advantages that access to mental models provides, thought experiments will be of use in the field of bioethics.

Having established that thought experiments are useful in bioethics, I then argued for the existence of literary thought experiments. I looked at a number of objections to the idea of literary thought experiments. I then considered two accounts of literary thought experiments, one by Carroll that is based in ethics, and one by Davenport that is founded on scientific thought experiments. To identify literature that is thought-experimental I look to McComb's family resemblance theory. Literature can be thought-experimental, and moreover the opportunity that literature provides to create extended narratives will in some cases be more useful for ethical understanding than philosophical thought experiments.

Science fiction has this same advantage: as ethical narratives it is detailed and humanized. In addition the speculative nature of science fiction lends itself to the exploration of new and emerging sciences and technologies including those in the field of bioethics. Thus science fiction thought experiments may indeed be better thought experiments for some emerging bioethical issues. Science fiction includes the elements of science, imagination and speculation, and this grasping of possibilities and use of alternative worlds fits with mental modeling theories of thought experiments. Utopias share characteristics and overlap with science fiction, and the particularly moral characteristic of utopias demonstrates clearly the ethical possibilities of literary and

science fiction thought experiments—that utopias, and literature generally, have narrative characteristics that enhance ethical understanding. Science fiction brings together science and ethics in powerful and informative thought experiments that are profoundly useful for bioethics.

Bibliography

Andler, Daniel. “The Undefinability of Analytic Philosophy.” Proceedings of the XXth World Congress of Philosophy, vol. 6, , Aki Kanimori, ed., Philosophy Documentation Center, Bowling Green State University, 2000, pp. 267-285.

Asaro, Catherine. "Complex speeds and special relativity." *American Journal of Physics* 64.4 (1996): 421.

Atwood, Margaret. *Oryx and Crake: a novel*. New York: Nan A. Talese, 2003.

Beck, Simon. "These Bizarre Fictions: Thought-experiments, our Psychology and our Selves." *Philosophical Papers* 35.1 (2012): 29-54. Print.

Berto, Francesco, "Impossible Worlds", *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/impossible-worlds/>.

Brison, Susan J.. "Outliving Oneself." *Aftermath: violence and the remaking of a self*. Princeton, N.J.: Princeton University Press, 2002. 12-39.

Brown, James Robert. *The laboratory of the mind: thought experiments in the natural sciences*. London: Routledge, 1991.

Brown, James Robert. "Why Thought Experiments Transcend Empiricism." *Contemporary debates in philosophy of science*. Ed. Christopher. Hitchcock, MA: Blackwell Pub., 2004. 23-43.

Brown, James Robert and Fehige, Yiftach, "Thought Experiments", *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2011/entries/thought-experiment/>.

Buzzoni, Marco. Empirical thought experiments: A trascendental-operational view. <http://philsci-archive.pitt.edu/4702/1/Empirical_Thought_Experiments._A_Trascendental-Operational_View.pdf>

Caput, Margherita. *Grazia Deledda's dance of modernity*. Toronto: University of Toronto Press, 2008.

Carroll, Noel. "The Wheel of Virtue: Art, Literature, and Moral Knowledge." *The Journal of Aesthetics and Art Criticism* 60.1 (2002): 3-26.

Cash, Mason. "Cognition without Borders: Socially Distributed Cognition, Non-Individualistic Autonomy and Moral Responsibility." Colloquium. Department of Philosophy Dalhousie University. Dalhousie University, Halifax. 8 June 2012. Speech.

Champagne, M.. "Stuck between Two Soup Cans: Limits of Rational Decision." Congress of the Humanities and Social Sciences. Canadian Philosophical Association. University of Waterloo, Waterloo. 28 May 2012. Address.

Cole, David, "The Chinese Room Argument", *The Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2009/entries/chinese-room/>.

Cooper, Rachel. "Thought Experiments." *Metaphilosophy* 36.3 (2005): 328-347.

Czerneda, Julie. "Science Fiction & Scientific Literacy." *The Science Teacher* 73.2 (2006): 38 -42.

Davenport, Edward. "Literature as Thought Experiment (On Aiding and Abetting the Muse)." *Philosophy of the Social Sciences* 13.3 (1983): 279-306.

Devlin, Patrick. *The Enforcement of Morals*. London: Oxford University Press, 1965.

Di Filippo, Paul . "A science-fiction fantasy." *Nature* 465 (2010): 1110.

Doan, Michael. "Famine, Affluence, and Morality." Justice in Global Perspective. Dalhousie. Dalhousie University, Halifax. 18 June 2012. Class lecture.

Donchin, Anne, "Feminist Bioethics", *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/fall2012/entries/feminist-bioethics/>.

Foot, Philippa. "The Problem of Abortion and the Doctrine of the Double Effect." *Virtues and vices and other essays in moral philosophy*. Berkeley: University of California Press, 1978. 20-33.

Fox, Renée C., and Judith P. Swazey. "“Hello, Dolly” Bioethics in the Media." *Observing bioethics*. Oxford: Oxford University Press, 2008. 85-128.

Frappier, Mélanie, Letitia Meynell, and James Robert Brown. "Introduction." *Thought Experiments in Philosophy, Science, and the Arts*. New York: Routledge, 2013. 1-11.

Galilei, Galileo, *Dialogue Concerning the Two Chief World Systems: Ptolemaic and Copernican.* Trans. with notes by Stillman Drake, New York: The Modern Library, 1953. 161-73.

Gatens, Moira. "The Art and Philosophy of George Eliot." *Philosophy and Literature* 33.1 (2009): 73-90

Gendler, Tamar. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium." *Midwest Studies In Philosophy* XXXI (2007): 68-89.

Goodenough, Jeremy. "The Trouble with Thought Experiments." *Theoretical & Applied Ethics* 1.2 (2011): 7-12.

Hayles, N. Katherine. *How we became posthuman: virtual bodies in cybernetics, literature, and informatics*. Chicago, Ill.: University of Chicago Press, 1999.

Hintikka, Jaakko. "The Emperor's New Intuitions." *The Journal of Philosophy* 96.3 (1999): 127-147. *JSTOR*. Web. 22 Apr. 2012.

Horowitz, Tamara, and Gerald J. Massey. *Thought Experiments in Science and*

Philosophy. Savage, Md.: Rowman & Littlefield, 1991.

Ichikawa, Jonathan, and Benjamin Jarvis. "Thought-experiment intuitions and truth in fiction." *Philos Stud* 142 (2009): 221–246.

Itzkoff, Dave. "It's All Geek to Me ." *The New York Times* 3 May 2006: n. pag. *The New York Times*. Web. 29 July 2012.

Jones, Tim. "Genre Benders: How Interstitial Fiction: Bringing Speculative and Literary Fiction Together." *English in Aotearoa* 67 (2009): 7-12.

Jorgensen, Darren. "The numerical verisimilitude of science fiction and EVE-Online." *Extrapolation* 51.1 (2010): 134-147.

Kirby, David A . "The Devil in Our DNA: A Brief History of Eugenics in Science Fiction Films." *Literature and Medicine* 26, no. 1 (2007): 83-110.

Knobe, J. (2003). “Intentional Action and Side Effects in Ordinary Language.” *Analysis,* 63, 190-193.

Kress, Nancy. *Beggars in Spain a novel*. New York: Avon Books, 1993.

Kress, Nancy. *Beggars & choosers*. New York: TOR, 1994.

Kress, Nancy. *Beggars ride*. New York: Tor, 1996.

Leri, Dennis. "Galileo." *SEMIOPHYSICS* . N.p., n.d. Web. 2 June 2012. <http://www.semiophysics.com/SemioPhysics_Articles_mental_3.html>.

Lufkin, Bryan. "Asteroid Mining in Fiction, Past and Present." *Innovationnewsdaily.com*. N.p., 23 Apr. 2012. Web. 17 July 2012. <http://www.innovationnewsdaily.com/1080-asteroid-mining-fiction-present.html>.

Marusek, David. *Counting heads*. New York: Tor, 2005.

Marusek, David. *Mind over ship*. New York: Tor, 2009.

McComb, Geordie. “Thought Experiment, Definition, and Literary Fiction.” *Thought Experiments in Philosophy, Science and the Arts*, eds., James Robert Brown, Mélanie Frappier and Letitia Meynell, Routledge: 207-222.

Miščević, Nenad. "Mental Models and Thought Experiments." *International Studies in the Philosophy of Science 6* 6.3 (1992): 215-226.

Miščević, Nenad. “Political Thought Experiments from Plato to Rawls.” *Thought Experiments in Philosophy, Science and the Arts*, eds., James Robert Brown, Mélanie Frappier and Letitia Meynell, Routledge: 191-206

Nersessian, Nancy J. "In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* Two: Symposia and Invited Papers (1992): 291-301.

Nilsen, E. S.. Graham, S. A., Smith, S. and Chambers, C. G. (2008), Preschoolers’ sensitivity to referential ambiguity: evidence for a dissociation between implicit understanding and explicit behavior. Developmental Science, 11: 556–562. doi: 10.1111/j.1467-7687.2008.00701.x

Norton, John D.."Chasing the Light: Einstein's Most Famous Thought Experiment," prepared for *Thought Experiments in Philosophy, Science and the Arts*, eds., James Robert Brown, Mélanie Frappier and Letitia Meynell, Routledge. (forthcoming): 123-140

Norton, John D.. "Einstein's Pathway." *University of Pittsburgh*. N.p., n.d. Web. 22 June 2012. <http://www.pitt.edu/~jdnorton/teaching/HPS_0410/chapters/general_relativity_pathway/index.html>.

Norton, John D.. "Why Thought Experiments do not Transcend Empiricism." *Contemporary debates in philosophy of science*. Ed. Christopher. Hitchcock, MA: Blackwell Pub., 2004. 23-43.

Perreault, Mathieu. "Des Mines en Orbite." *La Presse.* Montreal: 16 July 2012, sec. A: 14. Print.

Rawls, John. *A theory of justice*. Cambridge, MA: Belknap Press of Harvard University Press, 1971.

Shoulders, Felicity. "Conditional Love ." Asimov's Science Fiction, http://felicityshoulders.com/fiction/conditional-love/.

Silver, Lee M.. *Remaking Eden: cloning and beyond in a brave new world*. New York: Avon Books, 1997.

Stinner, Arthur. "Scientific Method, Imagination, and the Teaching of Physics." *La Physique au Canada* 59.6 (2003): 335-346.

Urquhart, Jane. "Intertextuality or Easter Egg Hunt: A Canadian writer's adventures in the library." Congress of the Humanities and Social Sciences. Congress of the Humanities and Social Sciences. University of Watertoo, Waterloo. 28 May 2012. Speech.

Worstall, Tim. "How Can the Sci Fi Billionaires' Planetary Resources Make a Profit From Mining Asteroids?" *Forbes.com*. N.p., 25 Apr. 2012. Web. 17 July 2012. <http://www.forbes.com/sites/timworstall/2012/04/25/how-planetary-resources-can-make-a-profit/>.

Associated Press in Washington. "Tech tycoons in asteroid mining venture." *The Guardian* . N.p., 24 Apr. 2012. Web. 17 July 2012. <http://www.guardian.co.uk/science/2012/apr/24/tech-tycoons-asteroid-mining-venture>.

"Space Innovators." Stewart, Jon . *The Daily Show*. Comedy Central. 25 Apr. 2012. Television.

"The Real Waterworld." *HistoricalAtlas.com: the Centennia Historical Atlas -- Europe and the Middle East* . N.p., n.d. Web. 23 May 2012. <http://www.clockwk.com/waterworld/index.html>