Thought Experiments & Literary Learning

by

Geordie Malcolm Alan M^{c}Comb

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Philosophy
University of Toronto

Abstract

Thought Experiments & Literary Learning

Geordie Malcolm Alan M^{c}Comb
Doctor of Philosophy
Graduate Department of Philosophy
University of Toronto
2020

In this work, I develop a novel approach to thought experiments and literary learning. It's novel primarily because, unlike many prominent approaches, it has us refrain from advancing theories, from giving logical analyses, and from explicating. We are, instead, to proceed in a way inspired by Wittgenstein's writings. We are, that is, to clarify words that give rise to problems and to clear those problems away. To clarify, we may compare language games in which figure terms like "thought experiment." Thereby, we might see that the concept these terms express has a family resemblance character. To clear away problems, we may describe how such a concept, if not illuminated, yields philosophical problems about thought experiments and literary learning.

After I develop this approach, I bring it to bear on two problems, and I achieve two main results. One problem concerns the nature of thought experiments. It is *Why do we have trouble explaining what we know them to be?* I find that, despite appearances, we have no such trouble. Central to this result are two claims about thought experiments. One is that imaginings aren't common to them. The other is that our unreflective concept of them has a family resemblance character.

The other problem concerns stories in works of literary fiction. It is *How could we possibly learn about the world from them*? To solve it, you might claim that we learn by performing thought experiments. And then you might appeal to a theory of them. I find that you'd risk explaining the wrong thing. That is, you may explain how we learn but not how we do so from literature itself. Central to this result are three claims, which concern how these stories and thought experiments differ. In short, they differ (i) in how we count imaginings as experiences of them, (ii) in how free we are to interpret them, and (iii) in how complex they may be. This last result achieved, I have twice taken my novel Wittgenstein-inspired approach, and there my dissertation ends.

Acknowledgements

I would like to express my appreciation to James R. Brown, my supervisor, for his longstanding and unflagging support. It began before I was a student at the University of Toronto, when he let me attend a workshop on thought experiments. Since then, by so many means—by mentorship and graduate seminar teaching, by conference and publishing opportunities, by preparing me for academic work, and by evaluating written and rewritten dissertation chapters—I've had his unwavering help.

I would also like to express my appreciation to Sonia Sedivy and Yiftach Fehige, members of my dissertation committee. I am especially grateful for their seeing my dissertation through, for their sharp criticism of the ideas in it, and for all that I have learned through my contact with them.

Thank you to my external examiners, Harald Wiltsche and William Seager, for their careful charitable reading of my work and for wonderful pressing questions.

Thank you as well to everyone who benefitted this work. They include Henry Laycock, who supervised my master's research project, as well as Joshua Mozersky and Sergio Sismondo, who evaluated it. They also include those who commented on this work's precursors, especially Michael Stuart, Mélanie Frappier, James Davies, and many of my fellow graduate students who were in my seminars or at my talks. Finally, they include members of my reading groups, especially Peter Alward, Erin Adrienne, Emer O'Hagan, Eric Dayton, Eran Tal, and Charles Repp.

The completion of my dissertation would not have been possible without the loving support of my wife, Erin. Thank you for everything you've done to help me finish this project, which took so much time so very often and spanned so much of our life together. Thank you too, Aidric and Eily, for giving me time to work even when you wanted me. And thank you, Mom, for hours of editing. Finally, I owe thanks to all my family, for time, for listening, and for all their loving help.

Contents

List of Figures

Preface

Dostoevsky's narrator, in *Demons*, asks Kirillov, "what, in your opinion, keeps people from suicide?"[1] Consider their ensuing discussion:

> "I… I still know little… Two prejudices keep them, two things; just two; one very small, the other very big. But the small one is also very big."
>
> "What is the small one?"
>
> "Pain."
>
> "Pain? Is it really so important… in this case?"
>
> "The foremost thing. There are two sorts: those who kill themselves from great sorrow, or anger, or the crazy ones, or whatever… They do it suddenly. They think little about pain and do it suddenly. But the ones who do it judiciously—they think a lot."
>
> "Are there any who do it judiciously?"
>
> "Very many. If it weren't for prejudice, there'd be more; very many; everybody."
>
> "Really? Everybody?"
>
> He did not reply.
>
> "But aren't there ways of dying without pain?"
>
> "Imagine," he stopped in front of me, "imagine a stone the size of a big house; it's hanging there, and you are under it; if it falls on you, on your head—will it be painful?"
>
> "A stone as big as a house? Naturally, it's frightening."
>
> "Fright is not the point; will it be painful?"
>
> "A stone as big as a mountain, millions of pounds? Of course, it wouldn't be painful at all."
>
> "But go and stand there in reality, and while it's hanging you'll be very much afraid of the pain. Every foremost scientist, foremost doctor, all, all of them will be very afraid. They'll all know it won't be painful, but they will all be very afraid it will be."

Kirillov invites the narrator to imagine a giant stone hanging over his head, then asks him whether or not being crushed by it would hurt. The narrator answers that obviously it wouldn't. Kirillov goes on to argue, from authority, that a judicious person would nevertheless fear that it would hurt—and, thereby, he explains why fear of pain is a prejudice that keeps judicious people from suicide. We may call this "Kirillov's giant stone thought experiment." Also, if we consider further scenes, such as Kirillov's suicide, we can interpret Dostoevsky as attempting to reveal a truth—e.g., that "apocalyptic intuitions and feelings," such as Kirillov's, if "divorced from a faith in Christ" and "turned into secular and subjective ideas," result in "monstrosities."[2] We may call this "Dostoevsky's Kirrilov thought experiment." To explain why we may so call them, I might point out similarities between such "literary thought experiments" and paradigmatic ones, such as Newton's Bucket and Thomson's Violinist. Then, to explain how we learn from these literary ones, I might appeal to theories of those paradigmatic ones.

1. Dostoevsky, *Demons: A Novel in Three Parts*, I.III.VIII.
2. Frank, *Dostoevsky: A Writer in his Time*, ch. 45.

With this in mind, ask yourself: How do thought experiments work? Are they like literary fictions? And what are they really? I will look into these and similar questions, and, to do so, I'd like to proceed as Roy Sorensen does. I'd like, that is, to give something like "a general theory of thought experiments: what they are; how they work; their virtues and vices."[3] For he promises his readers a familiar sort of account. But I won't. I'll even try not to advance a theory. Instead, I will look into such questions in Wittgenstein-inspired ways. Specifically, in what follows, I develop a line on thought experiments, which I extrapolate from Wittgenstein's writings, and which I extend to literary fiction. To elaborate, I'll characterize this line in three ways—in terms of problems, central claims, and contributions to the literature.

First, in terms of problems, I give two solutions. One problem is that we know what thought experiments are but find it hard to explain. My solution, that the hardness is only apparent, appeals primarily to a misunderstanding about definition but also to a "family resemblance" account of thought experiments. The other problem is that we cannot learn about the world from what isn't even about it, but we do so learn about it from stories in works of literary fiction, such as Dostoevsky's novels. My partial solution, that such learning may not be from literary fiction itself, appeals primarily to misunderstandings about imaginings in certain accounts of literary learning, i.e., in ones which have us either regard the stories as thought experiments or else equate the ones with the others.

Second, cast in the form of a chapter by chapter summary, here are my central claims. In the first, from Wittgenstein's writings, I claim to extrapolate three ideas: one is about how, by comparing "language games," we might shed light on our use of words like "thought experiment" and so clarify, e.g., the concept they express; another is about the form this concept takes, specifically, whether it has a "family resemblance" character; and, the third is about how, by clarifying words, we might, in a novel way, solve philosophical problems about thought experiments. These ideas, together, comprise an approach to questions like the above three.

This approach guides the second chapter, which has three parts. In the first, I clarify a question instead of defending an answer to it. The question is: What is a thought experiment? Asking it gives rise to a problem other than that of giving an answer. The problem has the form: How is our asking the question so much as possible? More specifically: How, given that we do recognize thought experiments, could we possibly have any trouble explaining what they are? I reject two answers. Then I clear away the problem. That is, I explain away the apparent trouble. To do so, I appeal to an illusion—i.e., that sharp definition alone can explain a nature—which prevents us from seeing how we already can and do explain what they are. This first part's overall claim, then, is that, once we see both the illusion and how we normally do answer our question about what they are, the problem dissolves. Then, in the second part, I defend this claim arguing for another, namely, that our unreflective concept of a thought experiment has a family resemblance character. To this end, I argue that no sharp definition captures our unreflective concept of them. The central premise is that they do not all have a certain "obviously essential property"—namely, imaginings. I also survey a swath of family resemblances that explain the concept, thereby helping us to see that they do; and, to give this survey some context, I stake out a position in the literature on when to give a "definition": never if sharply, now if non-stipulatively. In the third part, I explain why the claim about imaginings hardly rings true. To this end, I compare language games, aiming to describe an imaginings-free, stop-sign-like use of the expression "It's a thought experiment." The central contention is that, by overlooking this use, we may well doubt the imaginings-aren't-essential claim.

Finally, in my third chapter, I turn to works of literary fiction, especially various well-known novels in the

3. Sorensen, *Thought Experiments*, 3.

Western Cannon. Ask: How could we possibly learn about the world from stories in such works? My central claim, in short, is that, trying to solve this problem, and doing so in light of answers to strikingly similar ones about thought experiments, we risk losing our grip on something we want to explain. That is, we risk turning from our everyday ways of reading and reflecting to uses of the stories as thought experiments, which uses are either exceptional or ones we inadvertently invent. Ultimately this is so, I argue, because these uses differ from our everyday ones in three respects: in how imaginings function, in interpretational freedom, and in complexity. To illustrate the imaginings difference, if we read the above story from *Demons* as a literary thought experiment, we use our impending giant-stone imagining to feel *merely as a means* and not also as an end, unlike how we normally use it when appreciating the work. In so doing, one risks both losing one's grip on the story as literature ordinarily appreciated and, thereby, failing to explain how we learn from it. By shedding light on this risk, my aim is, in line with the above approach, to take a step toward clearing away misunderstandings and, with them, our problem about literary learning.

The third and final characterization I'll put in terms of my main contributions to the literature. Chapter 1 offers a novel Wittgenstein-inspired approach to the study of thought experiments, Chapter 2 a "family resemblance" account of our concept of a thought experiment,[4] and Chapter 3 an account of three epistemically significant differences between thought experiments and stories in works of literary fiction. Overall, the importance of these contributions lies primarily, I think, in how they work out ideas which run against the grain of certain others that are fundamental to the thought experiments literature.

4. I've given a similar but less developed account in McComb, "Thought Experiment, Definition, and Literary Fiction."

1 A Wittgenstein-Inspired Approach to Thought Experiments

In this chapter, from Wittgenstein's writings, I claim to extrapolate a new approach to questions like these: What are thought experiments? How do they work? Do literary fictions work similarly?

Let me explain this claim. To "extrapolate from Wittgenstein's writings" means to extend lines from his remarks, after I interpret them. The extended lines, taken together, outline my "approach" to questions like those above. "Like questions" include, beyond those listed, similar ones of course but, especially, those about words and concepts, e.g., about what concept we express using terms like "thought experiment." This approach to such questions, insofar as they amount to problems, is "new" in that it differs, as follows, from certain well-known others in the philosophical literature. Like certain of those in the analytic tradition, it has us—ultimately, by clarifying words—try to solve philosophical problems. Unlike them, however, we are neither to revise our usage nor to capture it with a fine-grained formal language. Unlike yet other approaches, to solve them, we are not to advance any theory.

This chapter comprises two sections. The first focuses on the clarifying, the second the novel problem solving.

1.1 A Way to Clarify Our Concept of a Thought Experiment

On the extrapolated approach, to clarify we are to compare. I explain this comparing in the first subsection. Its central claim is that, from Wittgenstein's writings, we can extrapolate a comparing use of "language games," one for clarifying, among other things, our concept of a thought experiment. This concept's form is the second subsection's subject. There, my central claim is that we can extrapolate a character the concept might have, namely, a "family resemblance" one. Altogether, in this section, I argue that we can extrapolate an approach to clarifying the concept, by comparing language games, in light of the possibility that it has that character. In the next chapter, I argue that the concept does indeed have that character.

1.1.1 Clarifying with Language Games

Again, the plan is, first, to interpret certain of Wittgenstein's remarks and then to extrapolate from them.

Interpretation Wittgenstein opens the *Philosophical Investigations* with a passage from Augustine's *Confessions*, one which gets across, it seems to him, an idea about the meaning of words—an idea which arises from a certain picture of human language; that is, he thinks, Augustine's picture of the essence of human language—that, roughly, our language is a bunch of names for things—gives rise to the idea that every word correlates with a meaning, i.e., the thing it stands for, and this idea, if not *itself* a primitive one about the way language functions, is *about* a language more primitive than our own.[1] Such a more primitive language, for

1. *PI* §1. All references to Wittgenstein's *Philosophical Investigations* are to the fourth edition (Wittgenstein, *Philosophische Untersuchungen: Philosophical Investigations*). In the footnotes, I'll flag salient translation differences with the third (Wittgenstein, *Philosophische Unter-*

example, is one "meant to serve for communication between a builder A and an assistant B," which Wittgenstein elaborates as follows:

> A is building with building stones: there are blocks, pillars, slabs and beams. B has to pass him the stones and to do so in the order in which A needs them. For this purpose they make use of a language consisting of the words "block," "pillar," "slab," "beam." A calls them out; B brings the stone which he has learnt to bring at such-and-such a call.[2]

This primitive language consists in four words, each of which names something—"block" a block, "pillar" a pillar, and so on—and every word has a meaning, i.e., the thing it's correlated with and for which it stands—the meaning of "block" being a block, that of "pillar" a pillar, and so on. To see that the language is primitive, *viz.*, simple relative to ours, notice that it lacks certain word types, such as demonstratives and proper names. These linguistic types, and this will matter shortly, Wittgenstein later builds into the language.[3] By analogy, *Peanuts'* Lucy might ask for *that* block or the block *named* "Charlie Brown," and English provides for the possibility of asking either of these two questions, but the primitive language does not, unless we build in further word types. Now, here is the point of all this. I've characterized the building-block language to prepare us for Wittgenstein's explanation of the term "language game," in which this language figures.

Here is the explanation. He calls, or might call, four sorts of activities "language games": first, what we can think of as the native-language learning games of children; second, primitive languages, such as the building-stone one; third, certain processes of naming and of repeating words, such as ring-a-ring-a-roses; and fourth, the whole made up of language and actions, the former "woven into" the latter.[4] This explanation, to be sure, names no commonality but, instead, points out certain activities, ones which seem variously similar. But it need not fail. For the concept explained, i.e., that of a language game, may well have a "family resemblance" character—on which, see below.

We've now glimpsed what language games are, and, before illustrating them at length, we will look at a use of theirs. The use on which we'll focus I'll describe first in general terms and then in relatively specific ones. As a slogan: language games model language to solve problems. That is, one way Wittgenstein uses them is to illuminate our language by comparing them to it, and this illumination helps solve philosophical problems. These illuminating uses, he describes, more specifically, as follows: "Our clear and simple language-games… stand there as *objects of comparison* which, through similarities and dissimilarities, are meant to throw light on features of our language."[5] We can see this comparing, for example, in how he "build[s] up the complicated forms [of our language] from the primitive ones by gradually adding new forms,"[6] e.g., by adding demonstratives, and then proper names, and so on, to the relatively clear and simple building-block language to shed light on, e.g., how our language outstrips Augustine's oversimplified idea of it. Now, these illuminating uses of language games somewhat resemble first approximations of physical phenomena, such as describing falling bodies but ignoring air resistance; and so—it might seem—they ultimately aim to help us give a physics-like, rigidly-systematic account of linguistic phenomena.[7] Alternately—again, it might seem—the language games so used aim to reveal preconceptions, or categories of the understanding, "to which reality

suchungen: Philosophical Investigations). Also in the footnotes, I'll refer to his *Tractatus* (TLP) (Wittgenstein, *Tractatus Logico Philosophicus*). I'll do so to help us see matters "in the right light," i.e., against the background of his "older way of thinking" (Wittgenstein, *Philosophische Untersuchungen: Philosophical Investigations*, 4e). For example, here I might flag *TLP* §3.203, where he had once been more favourably disposed toward this idea about meaning.

2. *PI* §2.
3. *PI* §8 & §15.
4. *PI* §7.
5. *PI* §130.
6. Wittgenstein, *The Blue and Brown Books*, 17.
7. *PI* §130.

must correspond."[8] But they have no such aims. For Wittgenstein aims, ultimately, to assemble and arrange what's illuminated about language to clear away philosophical problems—on which, more next section. Now, for a more specific construal of this comparing use, consider a dictum of Wittgenstein's, namely, that "if you want to understand the use of the word 'meaning', look for what are called 'explanations of meaning'."[9] This suggests that, if we want to shed light on the use of "meaning," we would do well to compare language games which include explanations of meaning, e.g., a teacher giving one. Alternately, we might consider such explanations while comparing.

To wrap up our interpretation of these remarks, and to bring out how general such uses of language games can be, I'll illustrate, in some detail, two such uses. Each sheds light on a different, albeit connected, feature of our language—one on our use of the word "meaning," the other on our rules for applying words.

For the first, take Wittgenstein's claim that—granted we may sometimes explain the meaning of a word by pointing at its bearer—for many uses of the word "meaning," we can explain it saying something of the form: "the meaning of a word is its use in the language."[10] For example, to explain the meaning of "pillar" in the building-stone language game, a builder in the game might point at a pillar and name it. Alternately, we, outside the game, might describe the word's use in that language, i.e., A's use of the word to get B to bring a pillar-shaped building stone as opposed to one of the others. Now, to account for how our description of the word's use, as opposed to some pointing, could explain its meaning in the language game, we might say that, in this case, as happens very often, what we'd call a word's "meaning" is its use in the language and not its bearer. Wittgenstein doesn't use this builder language game, as I've just done, to shed light on what he claims here about meaning and use. Rather, and this is the important thing, he does employ one. That he does so is, arguably, clear from the context.[11] Here, briefly, is the language game and its use. Suppose a sword named "Nothung" were shattered and ceased to be. The sentence "Nothung has a sharp blade" may well still have a meaning. But how could that be? After all, "Nothung" has no bearer, and, at least on Augustine's picture of language, that's the meaning. Well, he says, it still means something because "in this language-game a name is also used in the absence of its bearer."[12] We are to see, in this clear and simple language game and in how he accounts for the meaning of "Nothung," a certain difference—one between explaining a word's meaning by pointing at its bearer and doing so by describing its use. Having seen this difference there, we're supposed to be better able to make it out in our language; that is, we're to be in position to see the fact about our use of "meaning" in his claim about meaning and use. Seeing this fact, moreover, we are to gain an insight into a problem, similar to the Nothung one, in our language—i.e., "the problem of empty names"; that is, we are to see that, since, very often, we identify meaning with use instead of a bearer, there can be, as there obviously are, meaningful words without bearers, because words without them may nevertheless have a use. In sum, as I read these remarks, we are, by comparing the Nothung language game to our language, to shed light on a fact about our use of "meaning"; and, this comparing is to help us clear away a philosophical problem. Put another way, we are, by modelling linguistic meaning on this language game, to recognize his claim to be a fact—one among other facts that we are to order, aiming to solve philosophical problems, such as that of empty names—as opposed to systematizing language, e.g., establishing a theory on which meaning is use. Also, in light of the above claim's qualification, i.e., "for a large class of cases but not for all," the language game isn't supposed to exemplify a presupposition to which reality must correspond, e.g., that a satisfactory explanation of a word's meaning must appeal to its use in a given language.

8. *PI*§131.
9. *PI* §560; Cf. Wittgenstein, *The Blue and Brown Books*, 1.
10. *PI*§43.
11. *PI* §§39–44.
12. *PI* §44.

For the second illustration, ask: Could our rules for the application of words be gapless, i.e., determinate in every possible case? In particular, could we close every gap which arises from the possibility of applying rules differently? To shed light on the question, Wittgenstein modifies the above building-block language game.[13] First, he has us imagine builder A not calling but giving B written signs, which B interprets with a chart; on it are two columns, one of signs, the other building-block shapes, and the signs correspond one-to-one with shapes in the same row. This chart, moreover, is a rule B follows carrying out orders. Second, he has us introduce schemas for reading the chart and so different ways for B to read it. Instead of taking signs to correspond to a shape in the same row, for instance, a schema may tell B to take the sign to correspond to a shape one row down. These schemas are rules for chart reading, that is, rules for applying rules. They explain how to read the chart, which in turn explains what stone is to be brought at what written sign. Now, having so modified the building-block language game, Wittgenstein draws our attention to two possibilities. The first is of a regress, by asking whether or not we can imagine further rules to explain how to read the schema. The second is of there not being, despite any regresses, any need to close gaps, by asking whether the chart is incomplete without the schema, i.e., whether B can use it without any further rules that explain how to do so. Here is the upshot. By comparison with this clear and simple language game, he aims to clarify our obscure and complicated language in two respects and so solve a problem. That is, he tries to shed light on both how a regress may arise from differently applied rules and how such a regress might not be at all vicious, i.e., that we need not perhaps take its first step and, in so doing, take on an explanatory debt we can't repay—which, by the way, sets up his claim in the following remark that, under certain conditions, we need not take it on.[14] By contrast, this language game is neither a preparatory study for the future regimentation of our language, e.g., for laying down rules for applying rules, nor a preconception to which reality must correspond, e.g., a proof that, for any word, there's a regress of rules for its use.

Extrapolation From these remarks, as I interpret them, I'll now extend a line. It is this. To clarify words like "thought experiment": first, find or imagine clear and simple language games in which such words are used; second, compare them to our language; and, third, in so doing, look to explanations of what such words mean. But can we take such a line? I'll try to assuage four worries that we cannot.

First, are these words ever unclear? Yes, for example, we are sometimes at a loss asking ourselves what "thought experiment" means or what concept the term expresses. To be sure, if you're not now at a loss, and you explain the meaning—saying something like, "contemplation of an imaginary scenario," or "to conduct a thought experiment is to make a judgment about what would be the case if the particular state of affairs described in some imaginary scenario were actual,"[15] or else "basically devices of the imagination"[16]—then what exactly ought you to make of thought experiments in which you are supposed *not* to be able to imagine something, as in Nagel's Bat?[17]

Second, do such words have uses in our language? Yes. Some examples: we read and hear expressions of the form "So-and-so's such-and-such thought experiment" or "it's a thought experiment" and respond appropriately, i.e., read the paper that originated the thought experiment; or we write a response to a criticism of it; or we vary the imaginary situation summarized aiming to evaluate it; or we stop doubting that what's

13. *PI*§86.

14. Cf.: "The signpost is in order—if, under normal circumstances, it fulfils its purpose" (*PI* §87). Also, here I partly follow John McDowell, who claims: "Wittgenstein's regress shows that acting on an understanding cannot in general be acting on an interpretation of what is understood" (McDowell, "How Not to Read *Philosophical Investigation*: Brandom's Wittgenstein," 103).

15. Cooper, "Thought Experiments," 328–9; For this account, Cooper cites a well-known paper of Tamar Gendler's (Gendler, "Galileo and the Indispensability of Scientific Thought Experiment").

16. Brown and Fehige, "Thought Experiments."

17. For more on this, see §2.2.1.

described really happened.

Third, can we find or imagine uses of the words in clear and simple language games? Yes, in and around classes, for example, as we'll see next chapter, in §2.1.3. There, students regularly learn, in discussion or from teachers, to call class material "thought experiments," or the like, as opposed to "facts," or "reports," or "experiments," and so on—and, we can call some of these learning activities, or our imaginings of it, "language games." We can do so in at least two of the above four senses. First, we can think of certain of these activities as whole processes of using those words while learning them—as we do games by means of which children learn their native language. Second, we can think of the language in those learning activities as primitive, or simple like the above building-block game, relative to, e.g., that in the thought experiments literature.

Fourth and finally, how can we, comparing language games, look to explanations of what the words mean? That is, in light of the above dictum, how might we look for or imagine language games in which, or else outside of which, such explanations figure? We may begin by recalling familiar forms of explanation, such as giving examples or analyzing. That is, first, often, to explain what the words mean, we, as it were, point at their bearers; for instance, we either point out the paper in which a passage expresses so-and-so's such-and-such thought experiment, or else we remind someone of one by referring to its distinctive striking imaginary situation or event. Alternately, second, we often, as it were, break the term into its component parts and then arrange the result; for example, we try—assuming the term consists in an adjective modifying a noun—to give a synonym, such as "experiment in thought," "mental test," or, trying to be more precise, "testing a theory by manipulating hypothetical variables and observing the result." Also, we often do both, i.e., try to explain the meaning with an analysis and fitting examples. Finally, we may marry these familiar forms of explanation to language games used as above. That is, we may (i) try to recall such familiar by-explanation-or-analysis forms of explanation, e.g., ones in and around classes or in the philosophical literature, and then (ii) find or imagine clear and simple language games in which, or else outside of which, these forms figure, and then, finally, (iii) compare these games to our language to clarify words like "thought experiment."

1.1.2 Family Resemblance Concepts

So far, we have extrapolated a method, that is, a use of language games for clarifying words like "thought experiment," e.g., for shedding light on their meaning or the rules for their application. We might also, by appeal to clarified usage or rules, try to clarify the concept expressed. To do so isn't foreign to Wittgenstein's writings, since he clarifies concepts by appeal to such things, e.g., that of a game by appeal to what can be so called, as we'll shortly see. More to the point, so using the method, we may find that the concept has a "family resemblance" character. In what follows, I'll extrapolate the idea that it does.

Interpretation Against the above explanation of language games, and what he says about them, Wittgenstein raises the objection, in short, that he, being lazy, failed to say what the essence of language is.[18] He replies that he hasn't failed. In particular, he says he hasn't tried, as might seem required, to find the property common to all of language that accounts for our calling it "language." But he also doesn't think such a property exists. Instead, he thinks, in short, that "affinities" do the job.[19] Evidently, this rules out neither that commonalities exist nor that they jointly explain.[20] That is, for all Wittgenstein says here, he allows that

18. *PI* §65.
19. The third edition has "relations" and cognates in place of the richer term "affinity" and its cognates.
20. To interpret otherwise, as some do, argues Michael Forster, is to misinterpret (Forster, "Wittgenstein on Family Resemblance Concepts," 69).

what's called "language" might have commonalities, such as being arbitrary or artificial, and be explained by them together with affinities.

So far, although affinities and commonalities play the same explanatory role, they differ. To explain how, Wittgenstein, among other things, draws our attention to various groups of activities called "games." If we look at them, instead of merely thinking, he says, we will see no commonality but instead similarities—specifically, "a complicated network of similarities overlapping and criss-crossing: similarities in the large and in the small."[21] The affinities are such similarities. Now, if affinities among games turn out not to explain why we call them so,[22] then Wittgenstein hasn't shown the phenomena to exist, but even then he has nevertheless shed light on what affinities are.

Then, in effect, he coins terms. He calls the affinities "family resemblances"—since, he says, they criss-cross and overlap as do the relations between family members, with respect to "build, features, colour of eyes, gait, temperament, and so on"—and, he adds, to say games stand in these relations, he'll say that they "form a family."[23]

But how do overlapping and criss-crossing differ? What is it to be both? And how might large and small similarities fit in? Well, orderly overlapping looks like this:

> Game A might be a game in virtue of being a, b, c; B one in virtue of b, c, d; and, C in virtue of c, d, a.

Disorderly overlapping looks like this:

> [G]ame A might be a game in virtue of being a, b, c; B one in virtue of b, c, d; and C one in virtue of d, e, a; D one in virtue of e, f, g; and so on.[24]

Criss-crossing is a disorderly kind of overlapping. It isn't, as Michael Forster would have it, orderly but only slightly overlapping similarities:

> [G]ame A might be a game in virtue of having features a, b, and c; game B in virtue of having features a, d and e; game C in virtue of having features d, f and g; and so on.[25]

By "overlapping and criss-crossing," then, Wittgenstein means not orderly but a kind of disorderly overlapping. But then why didn't he make it explicit? It arguably already is. We often read "overlapping" as "orderly overlapping," as we often read "family" as "non-extended family," and the addition of "criss-crossing" in the *PI*, a term absent in the earlier *Blue Book*,[26] prevents this too narrow reading by drawing our attention to a kind of overlapping we're inclined to overlook. By analogy, we can, by adding the expression "and extended-family" to "family," clearly mean family broadly construed. Finally, similarities that overlap and criss-cross *in the large and in the small* look like the above criss-crossing after we specify that features *a, b, c…* vary from details to general features, e.g., from numbers of dice to the purposes of rolling them.

These touchstone remarks now summarized, we can say:

> Affinity-explained words express family resemblance concepts.

That is, a concept has a family resemblance character if (i) the members of its extension stand in family resemblance relations—which are, paradigmatically, overlapping and criss-crossing similarities, ones at various levels of generality—(ii) the words that express it have a calling use, and (iii) these resemblances, as opposed to

21. The third edition offers a narrower translation: "sometimes overall similarities, sometimes similarities of detail."
22. Cf. Suits, *The Grasshopper: Games, Life and Utopia*.
23. *PI* §67.
24. Forster, "Wittgenstein on Family Resemblance Concepts," 71.
25. Forster, 71.
26. Wittgenstein, *The Blue and Brown Books*, 17.

a lone commonality, account for that use. For an alternative characterization, let us, first, abstract from explaining this calling use to explaining concept application; and, second, let us, following Forster, distinguish between explaining applications and defining the concept.[27] Doing so, we can say:

> Concepts we apply in virtue of affinities and cannot define are family resemblance ones.

That is, a concept has a family resemblance character if (i) we apply it by means of family resemblances instantiated in its extension, as opposed to a lone commonality, and (ii) we cannot explain it by appeal to such a commonality alone. On this, three points. First, it evidently doesn't follow that all concepts have a family resemblance character, and arguably it isn't so for Wittgenstein.[28] Second, "cannot define" here means cannot explain by appeal to a lone commonality, e.g., a set of individually necessary and jointly sufficient conditions.[29] By contrast, and this will matter shortly, I don't take it to mean: cannot explain by appeal to certain similarities, e.g., to a set of overlapping disjoint conditions not one of which is necessary but each of which is sufficient. Third, the characterization does not sharply delimit the class of concepts that have a family resemblance character. To illustrate, take concept *F*, which applies solely in virtue of a few minimally overlapping similarities, e.g., being *ab*, *bc*, or *cd*. *F* has a family resemblance character only if these similarities amount to affinities—but, so far as the above characterization goes, whether or not they so amount is indeterminate.

Wittgenstein then develops his characterization. To do so—changing his example from games to numbers and, part-way through, his analogy from family resemblances to fibres in a thread—he describes how a certain family resemblance concept changes and persists.[30] I divide his remarks in three: first, kinds of number form a family, and we might call something "a number" if it's like, has a "direct affinity" with, some members of the number family—even if it's also unlike others, with which it has an "indirect affinity"; second, calling, in this way, what hasn't yet been called "number," we, like adding fibres to lengthen a thread, lengthen our concept of a number; and, third, it is affinities, or fibres, and their overlapping in particular—as opposed to a commonality—that explains the thread's "strength." That is, it explains our confidence that we possess the concept.[31] For example, the overlap explains our confidence that we apply one and the same concept of number, as opposed to a plurality of distinct ones, over time to new and different objects.[32]

After developing his characterization in this way, Wittgenstein considers a definition. On it, a family resemblance concept is a cluster of overlapping sub-concepts. That of number, for instance, is to be "explained as the logical sum of those individual interrelated concepts: cardinal numbers, rational numbers, real numbers, and so forth."[33] He takes issue with this definition, specifically with the presumption that such concepts *must* be so explained. His reason is that, although we do use words like "number" for such "unbounded" concepts, we could instead use them for "rigidly bounded" ones were we to draw them such a boundary, e.g., establish that we are to call all and only what shares such-and-such set of features "a number." In short, to advance the definition is to overlook sharp stipulation.

27. Forster, "Wittgenstein on Family Resemblance Concepts," 73. Contra Forster, Wittgenstein, I think, presupposes this distinction but doesn't *fail* to draw it.

28. His considered position is, arguably, that some but not "*all* general concepts work this way" (Forster, 67).

29. Forster further requires that such conditions be "non-trivial" and "essential," i.e., to provide an analysis and to explain the nature of. For example, holiness doesn't define piety, even if it's a necessary and sufficient condition, because it's trivial, and to be loved by the gods doesn't either because it isn't essential (Forster, 71 n. 25).

30. *PI* §67.

31. See Wittgenstein's explanation of "strength" in Klagge, "The Wittgenstein Lectures," 363–364.

32. Cf. Hacker's claim that new members of a family resemblance concept's extension "accrue or can accrue without any change in the concept" (Baker and Hacker, *Wittgenstein, Understanding and Meaning: Volume 1 of an Analytic Commentary on the Philosophical Investigations, Part II*, 171). This claim alone, to be sure, clearly isn't "essentialist" about such concepts. Klagge may mistakenly think otherwise (Klagge, "Wittgenstein and von Wright on Goodness," 295).

33. *PI*§68. The third edition has "defined" instead of "explained."

But isn't the fault he points out, rather, that a definition is given at all—on the grounds that family resemblance concepts are indefinable? Arguably not. First, he doesn't deny that a definition, such as the "unbounded" one of number above, could capture a family resemblance concept—only that it must, because of stipulation. Second, he doesn't deny that any definition whatsoever can explain a family resemblance concept. That is, as we characterized family resemblance concepts above, they're not definable insofar as we cannot explain them by appeal to a lone common property—but this doesn't rule out the possibility of defining them in general. After all, we may nevertheless appeal, e.g., to a set of unbounded disjunctive conditions like those in the number definition above. To be clear, my point here concerns what is ruled out, not whether another sort of definition, e.g., a disjunctive one of number, is really possible. Let me add to this that, for Wittgenstein, such a disjunctive set isn't itself a common property, since, he says, to say so is to play with words.[34] But, then, if we so read him, don't we fail to appreciate "some of the most philosophically interesting implications of the phenomenon" of family resemblance concepts, as Forster argues?[35] To illustrate such an implication, if family resemblance concepts aren't *in any way* definable—if no such rule governs them—and all concepts must be rule-governed, then how could they even count as concepts? This is one problem, but not the only one that can count as a philosophically interesting implication. Another, for instance, is a specific version of it. Namely, how could there be family resemblance concepts, which cannot be defined by appeal to a lone common property, if concepts in general must be rule-governed? Now, in light of the seemingly easy solution that disjunctive conditions could form such a rule, Forster may reply that "the problem's full force" requires that the concepts be indefinable even with such conditions.[36] But, if so, it's no longer clear that the problem is one that Wittgenstein addresses or is, therefore, relevant here.

How, we may now ask, do we explain family resemblance concepts? Sometimes, with examples and, perhaps, rules for their use—as it were, with base case and inductive step. That is, Wittgenstein thinks, to explain, e.g., what a game is, we "describe *games*" and might add "'This *and similar things* are called 'games'."[37] If added, this rule doesn't specify a commonality but, instead, various similarities, which could, in principle, be the affinities in virtue of which we apply the relevant concept. Alternately, without the rule's addition, the described examples alone could, along the same lines, specify those affinities. In this way, one may come to know how to work out what falls under the concept, i.e., come to possess it. And such an explanation may be complete, not a mere approximation of some "unformulated definition" in our minds, which determines the concept's extension.[38]

One last interpretive point. These remarks suggest that, for Wittgenstein, family resemblance concepts are vague, i.e., have indeterminate extensions.[39] Isn't this to conflate matters, i.e., family resemblance with vagueness, as Forster claims?[40] Arguably not. First, Forster claims that family resemblance concepts, as sketched above, could have determinate extensions, even if in practice none do. But that's questionable. For sub-concepts as such must, it seems, disagree over whether the concept comprising them applies in some possible case. For example, before the establishment of imaginary numbers as such, sub-concepts would presumably have disagreed over whether our concept of number applies to "$\sqrt{-1}$." And, in such cases, extensions would have to be indeterminate. Still, to be necessary isn't to be essential. That is, we can explain family resemblance without explaining vagueness, as we can explain 2 without explaining $\sqrt{4}$. Here, then, is

34. *PI* §67.
35. Forster, "Wittgenstein on Family Resemblance Concepts," 74.
36. Cf. Forster, 76.
37. *PI* §69.
38. Cf. *PI* §75.
39. Cf. Von Wright's claim, charitably read as Klagge reads it, that a characteristic of these concepts is "Bewilderment as to whether something 'really' falls under" them (Wright, *Varieties of Goodness*, 16 & Klagge, "Wittgenstein and von Wright on Goodness," 294).
40. Forster, "Wittgenstein on Family Resemblance Concepts," 70.

the second point. There's no conflation if we do not read these remarks, as we did not read Wittgenstein's thread ones above, as aiming solely to explain the nature of family resemblance concepts. And we need not so read them. We may, instead, read them as aiming to describe some actual explanations of the concepts—which, in principle, may shed light on their nature without any conflation.

To sum up, here is the reading for which I've argued. A concept has a family resemblance character if (i) it applies in virtue of affinities—i.e., paradigmatically, a network of similarities overlapping and criss-crossing at various levels of generality—and (ii) it cannot be defined, i.e., stipulation aside, cannot be explained by appeal to a lone common property. Also, first, such concepts persist via overlapping similarities, which may increase, extending them. Second, we can and do explain them by giving examples, possibly alongside and-so-on rules. And, third, they're vague, although we explain family resemblance differently than vagueness.

Extrapolation Again, the idea I want to extrapolate is that our concept of a thought experiment has a family resemblance character. To do so, I won't simply infer that this concept might have that character from the propositions that "thought experiment" is a general term and that some such terms, like "game," might have it. I won't because the idea should, in light of its role in the extrapolated approach, be a real possibility—i.e., one plausible enough that we shouldn't dismiss it out of hand. To extrapolate, then, I'll argue that some words like "thought experiment" *may well* express family resemblance concepts—and I'll base this argument on two different premises. The main one is that the above touchstone sketch may well picture the relevant concept. That is, stipulation aside, affinities, instead of commonalities alone, may well sometimes explain why we call something a "thought experiment," or a like term—and so our concept of a thought experiment may well be a family resemblance one, not one that is either applied in virtue of a lone commonality or definable by means of a conjunctive set of necessary and sufficient conditions. The other premise, a subsidiary one, corroborates the preceding argument. It is that our concept of a thought experiment has four trappings of family resemblance concepts: first, that we can, by looking at how we use terms like "thought experiment," shed light on the concept; second, that we extend and strengthen it "twisting fibre upon fibre"; third, that we explain it with examples plus perhaps a rule; and, fourth, that it is vague.

In support of the first premise, when we look at what we call "thought experiments," and pay attention to what we'd appeal to were we explaining such calling, we see similarities crop up and disappear. For example, Galileo's falling bodies thought experiment has a certain outcome, but the Clock in a Box one does not, only one or another;[41] Thomson's Violinist and Newton's Bucket contain multiple hypothetical events, but Black's two spheres one has no events, only a situation;[42] Jackson's Mary the Colour Scientist and Einstein's elevator thought experiments have us imagine something but, again, Nagel's Bat has us fail to do so.[43] That is, since, at first pass, that in virtue of which we would explain what thought experiments are turns out, upon examination, merely to be similarities that don't amount to commonalities, we've good reason to think it plausible that (i) no commonality lies in them that explains our calling them "thought experiments" and (ii) similarities instead do the job; in turn, we've good reason to think, assuming the similarities may well amount to family resemblances, that, stipulation aside, our concept of a thought experiment may well be a family resemblance one—i.e., may well both apply in virtue of affinities and not be definable by means of a set of conjunctive necessary and sufficient conditions.

In support of the second, subsidiary premise, that our concept of a thought experiment has four trappings of family resemblance concepts, consider each of the four in turn. First, last paragraph, by looking at what we

41. For more on this difference, see §3.4.1.
42. For a description of this last case, see §2.3.1. For the first, see §2.3.2.
43. For more on this difference, see §2.2.1.

call a "thought experiment," as by looking at what we call "games," we saw, arguably, certain ways in which we use a term and, thereby, shed light on the concept it expresses. We'll see these ways again next chapter. This light shedding by examining usage is, arguably, one trapping of family resemblance concepts that our concept of a thought experiment has. The second trapping, that we extend and maintain our concept of a thought experiment "twisting fibre on fibre," we see in the case of whether or not to call certain works of literary fiction "thought experiments." For example, we sometimes call Orwell's *1984* one because, we think, it aims to establish a certain thesis by having us contemplate certain imaginary events, like many typical thought experiments, despite various dissimilarities with them, such as its complexity; and, thereby, if we're right, we extend the concept, allowing it to persist, instead of replacing it with a new one. I'll return to this topic in the final chapter. The third corroborative trapping is that we do sometimes learn the concept by means of examples plus perhaps a rule for identifying other members of its extension. Think, for instance, of how we learn from examples given in class and from explanations that run primarily on examples. Again, I elaborate next chapter. The fourth and final trapping is, plausibly, that the concept is vague—as family resemblance concepts aren't essentially but, as I argued above, must be.

Finally, consider two worries. One is that, unlike the word "game," the term "thought experiment" isn't an everyday one; rather, it's formal or technical. Can such a term express a family resemblance concept? Perhaps highly technical terms cannot express them, e.g., ones used strictly in accord with an explicit sharp definition, e.g., that of "set" in set theory. But we hardly ever, e.g., in science, philosophy or the culture at large, use "thought experiment" strictly in accord with such a definition. Second, you can tell that I lack the genuine set-theoretic concept of a set, as opposed to our cutlery-related one, if you see me identify as the same two sets with different members—but how could you possibly tell whether or not I have a genuine concept of a thought experiment as opposed to one I've made up? After all, if it's a family resemblance one, and if we haven't agreed on a stipulation, there's no definition like that of "set" on which you could rely.[44] Well, you may ask me to give examples and, perhaps, a similarity rule, to explain what "thought experiment" means, and you could then, in principle, judge what concept I have based upon such explanations. After all, I take it, for Wittgenstein, to master a use of words, or to know what they mean, is to understand them, or to know what concept they express;[45] and we can judge whether or not someone has mastered the use, e.g., in how that person explains the term.

1.2 A Novel Approach to Philosophical Problems about Thought Experiments

So far, I've extrapolated a way to clarify words like "thought experiment." This includes what the words mean, the rules for their use, and the concept they express. To so clarify, we are to compare our language to real or imagined language games in which words like "thought experiment" have a use—all with an eye both to explanations of what the words mean and to the possibility that our concept of a thought experiment has a family resemblance character. In what follows, first, I'll extrapolate an approach that runs on clarifying words, e.g., by comparing language games. The approach's aim is to clear away problems about thought experiments, e.g., those about what thought experiments are, how they work, and whether they work like narrative literary fictions. To do so, the approach has us give two accounts, one of how misunderstanding terms like "thought experiment" leads us into problems, the other of how by clarifying these terms we can solve those problems.

44. Cf. Forster's concern about appeals to family resemblances to justify classificatory whims (Forster, "Wittgenstein on Family Resemblance Concepts," 83).

45. Here, I follow Hans-Johann Glock: "one can find, more or less explicitly, in Wittgenstein's later oeuvre… [the view that] Concept-possession is a particular kind of ability. To possess a concept is to have mastered the use of an expression" (Glock, "Wittgenstein on Concepts," 93).

Then, second, to end the chapter, I'll differentiate this approach from those in certain major positions on thought experiments. The differences, in short, are that it neither makes theoretical posits, nor reforms language, nor gives logical analyses.

1.2.1 The Problem Solving Approach

As above, before extrapolating, I interpret. My main interpretive goal is to bring out of Wittgenstein's writings an approach to problem solving that runs on clarifying words. The one I bring out consists in trying to give two accounts, one of how certain philosophical problems arise from our forms of expression, the other of how, by clarifying such forms, we clear away the problems. To bring out this approach, I examine three cases, the last in some detail. The first concerns solving a problem by clarifying those forms, the second how certain forms yield a problem, and the third how others do so. Then, interpretation done, I use these three cases as models to extrapolate an approach to thought experiments.

Interpretation Recall that, for Wittgenstein, from the Augustinian picture of the essence of human language, an idea arises, namely, that every word correlates with a meaning, i.e., the thing for which it stands. But then how could a name without a bearer have a meaning? To shed light on such a question, he considered a language game—one in which an expression, "Nothung has a sharp blade," has a use even though a name in it, i.e., "Nothung," has no bearer. The goal was to bring out how, in our language, very often, the meaning of a word, such as "Nothung," can be explained by describing its use, as opposed to pointing at its bearer. And a point of bringing this fact out, as we saw, was to shed light on the above question about empty names. For it helps us to see, e.g., that not every name, to mean something, must have a bearer and that we need not refer to a bearer to explain what an empty name means. Now, if successful, this clarification helps to clear problems away, since it removes certain underlying motivations, i.e., ones for positing a bearer to explain the sense of empty names. We might posit, to take an example not known to Wittgenstein, a bearer that doesn't exist in the actual world but does exist in another possible one, specifically, in one that is concrete and spatiotemporally distinct from our own, as David Lewis had it.[46] Without the motivation to make such a posit, certain problems fall away. After all, if we do not posit a possible bearer, we need not, in this particular case, work out whether or not these bearers exist in such possible worlds. In that case, questions like the following, for example, need not pose problems but may instead run idly: Were we to posit such worlds, might we violate certain scientific methodological principles—given that, in so doing, although we do not commit ourselves to the existence of any new kind of entity, we nevertheless commit ourselves to new entities of the same kind?[47]

Divide this first case into the above two sorts of account. The problem genesis one is of how a picture of language connects us to an idea, which idea requires some explanation, and which explanation raises philosophical problems. The other account, the problem solving one, is of how, by clarifying with a language game, we see the initial connection, which frees us from the explanatory requirement, which in turn prevents the problems from arising—at least from that source.

The second case concerns, primarily, an account of the problem genesis sort. It begins, in *PI* §93, with the claim that "the forms of the expressions we use in talking about propositions and thought" sometimes make us unable to see what's in plain view about propositions, in particular, how they work. Wittgenstein goes on to say that, first, by being unable to see it, and so misunderstanding our language's logic or rules, and that,

46. Cf. Lewis's counterfactual analyses of fictional statements, e.g., about Sherlock Holmes, for which he cites *Counterfactuals*, where he famously defends this view of possible worlds (Lewis, *Counterfactuals*, 42 ff.).

47. For Lewis' contention that a belief in the existence of possible worlds is *not* an ontologically unparsimonious hypothesis, see Lewis, 87. For replies, see Daniel, "Quantitative Parsimony" and Baker, "Quantatative Parsimony and Explanatory Power."

second, by recognizing the importance attaching to propositions, we are tempted into thinking that propositions behave extraordinarily, or even uniquely.[48] One such expression, from *PI* §92, is the question, "What is a proposition?"—insofar as the form we take it to have is "The essence is hidden from us." For then it renders us unable to see the in-plain-view workings of propositions. That is, under certain circumstance, if we take the expression to have that form, we're led away from seeing what's in plain view—namely, the function and structure of propositions—and toward an analysis, one aimed at digging something out of the phenomena, namely, their hidden essence.[49] Instead of such digging, he adds, we are (i) to think of the "essence" of language as something that is in plain view and (ii) to order its "structure, function" such that it becomes "surveyable." Finally, this problem genesis account figures in a problem solving one. That is, in light of *PI* §90, giving such an account of misleading forms of expression furthers an inquiry, one that aims to solve problems by clearing away misunderstandings about words.

To see in more detail how such problems may arise, let us turn to the third and final case. In *PI* §§88–89, following a remark about the adequacy of "inexact" expressions, such as "stay roughly here," Wittgenstein raises a problem, namely, "How is logic something sublime?" His treatment of the problem includes an account of how, misunderstanding "ideals," we think logic sublime. I'll now explain this account. Then I'll consider how it fits into a problem genesis account.

To explain such ideals, in *PI* §100, Wittgenstein has us consider a denial, namely, that something is not a game "if there is some vagueness *in the rules*"; and, he says, someone who makes it, if challenged, may concede, "Well, perhaps you'll call it a game, but at any rate it isn't a perfect game."[50] Such a "perfect game," Wittgenstein calls "an ideal," and such an ideal, he adds, may dazzle and confuse us. For example, we may be distracted by our ideal for the use of the word "game" and, misunderstanding its role, become confused like the denier. That is, we would call something with vague rules "a game," were we not "dazzled by the ideal" and, so distracted, unable "to see that actual application of the word 'game' clearly." Finally, to help prepare a place for such dazzling in a problem genesis account, notice how it differs from the above importance of propositions. Both help explain how, from a linguistic misunderstanding, a confused claim arises; however, whereas the dazzle helps by explaining this misunderstanding, the importance does not, instead helping alongside it.

Having introduced such ideals, Wittgenstein turns to his account's central move. In particular, he turns to a misunderstanding of an ideal that makes it appear necessarily real. He begins, in *PI* §101, with a tempting claim, one like the above denial, about logic's sublimity—namely, that "there can't be any vagueness in logic"—but, unlike the denial, he doesn't say that we misapply a term because of a dazzling ideal. Rather, he describes what happens next. Specifically, he describes how a certain idea now absorbs us, i.e., how it then forces itself upon us. The idea is that the ideal—i.e., vagueness-free logic—"'*must*' occur in reality." He goes on to characterize this along two lines. First, "one doesn't as yet see *how* it occurs there." Second, one "doesn't understand the nature of this 'must'." Then he explains how we, while neither seeing the one nor understanding the other, come to think that this ideal has to occur in reality. To do so, he says, "we think we already see it there." To sum up this central move, in light of the claim that logic has to be vagueness-free, which claim resembles the above confused one made about games when dazzled by their ideal, we're strongly inclined to think that a certain ideal—i.e., vagueness-free logic—"must" occur in reality, despite not seeing *the way in which* it occurs there and, all the while, not understanding what the word "must" means; and, to explain why we're so led, despite not seeing this and all the while not understanding that, he appeals to our thinking

48. Cf.: "The proposition constructs a world with the help of a logical scaffolding…" (*TLP* §4.023).

49. Cf. *TLP* §§3.34–3.341, where Wittgenstein tries to dig through accidental features of propositions down to the essential ones, and this leads him to the following picture, in *TLP* §4.5: "The general form of a proposition [i.e., that alone which is essential to it] is: Such and such is the case."

50. The third edition has "complete game" in place of "perfect game."

we already see the ideal in reality.

This explanation he then elaborates in two complementary ways. First, he uses an example to illustrate how, while not seeing how a certain ideal occurs in reality, we nevertheless come to think we see it there. Second, he uses metaphors and a simile to capture how, while misunderstanding what that "must" means, we think it "must" occur in reality. Consider each in turn.

The first elaboration's example, found in *PI* §102, has three parts: an ideal—*viz.*, the "strict and clear rules for the logical construction of a proposition"; thinking we see the rules in reality—*viz.*, these rules appearing to us as "something in the background—hidden in the medium of understanding"; and, doing so despite not seeing how they occur in reality—*viz.*, thinking we "see them (even though through a medium)." Then, to explain why, not seeing them, we aren't led to doubt that we see them, he appeals to our thinking we "understand the sign" or "mean something by it." For example, thinking we understand the expression "Ludy is Austrian and Bertie is English," we wouldn't doubt that we perceive, somehow in the medium of our understanding, a logical rule in virtue of which we think we understand it—e.g., a conjunction construction rule, such as $(p) \wedge (q) \rightarrow (p \wedge q)$—despite not seeing how we perceive it.

The second elaboration's metaphors and simile, found in *PI* §103, complement each other. The metaphors capture a picture we have of the necessity of an ideal.[51] It's "unshakable," something we "can't step outside of," something from which we "must always turn back," and something outside of which nothing is and in which we "cannot breathe." Then, the simile, by placing this picture of necessity in a broader landscape, helps us to see it in the right light. That is, the ideal, he goes on to say, "is like a pair of glasses on our nose through which we see whatever we look at. It never occurs to us to take them off." For example, a logical rule "must" be real, or its reality is "unshakeable" and so on, insofar as we do not think to put the rule aside. It is this aspect of an ideal's necessity that we misunderstand when thinking the ideal "must" occur in reality.

The next remark, *PI* §104, reaps what the last few sowed.[52] It consists of two sentences. The first is that one "predicates of the thing what lies in the mode of representation." That is, you do such things as use a general term to say (predicate) of what you're talking about that it is thus-and-so—and, the frame, or ideal, that you use to say it (your mode of representation) is itself thus-and-so. We might say, for example, that games or logical rules are vagueness-free, saying of them what lies in an ideal we use to talk about them. By the way, as we saw, confused, we may say such things because, dazzled by the ideal, we're distracted from actual usage, e.g., from the in-plain-view structure and function of words like "game" and "logic." Moving on, the second sentence concerns thinking we perceive an ideal in reality. It is that we "take the possibility of comparison, which impresses us, as the perception of a highly general state of affairs."[53] That is, first, we're impressed how well we can compare what we're talking about (representing) to the frame or ideal we're using to talk about it (to our mode of representation). For example, we're impressed how well a formal conjunction construction rule compares to an everyday inference in English such as, "Obviously, since each is a philosopher, both Bertie and Ludwig are." Second, we take this impressive possibility of comparison—e.g., of making out how well one matches the other—to be perceiving a very broad truth about the nature of what we're representing—as it were, seeing its general outline. By analogy, being awed by how well ideal construction rules line up with how we can build up English sentences, we might think we are perceiving a highly general truth, e.g., that those rules, somehow, lie within and structure those sentences. By the way, as we saw, via such an apparent perception of an ideal, e.g., of a logical rule lying somehow in the understanding's medium, we're led to think it "must" exist

51. One found, e.g., in *TLP* §4.12.
52. Cf. Wittgenstein, *Culture and Value*, 78e.
53. The third edition differs insofar as it has our being impressed lead to the mistake and has the state of affairs perceived be "of the highest generality."

in reality. Also, we're so led both despite our not seeing how it occurs there and without our understanding the necessity, that is, as it were, how its unshakeable-ness depends on our not thinking to remove our glasses.

So far, in this third case, we've been concerned with how, misunderstanding ideals, logic appears sublime. Let us turn to how some such appearances yield philosophical problems—in particular, how they yield two of their characteristics, namely, disquiet and persistence. Once we explain this yield, we'll have, if only partly and in outline, a problem genesis account. Consider each characteristic in turn.

In the last remark, we take a possibility of comparison to be a perception. Similarly, in *PI* §112, a "simile that has been absorbed into the forms of our language produces a false appearance which disquiets us." To characterize this disquiet, Wittgenstein, among other things, describes the kind of statements we might make when so appeared to, as follows: "'But *this* isn't how it is!'—we say. 'Yet *this* is how it *has to be*!"' For instance, if we think we perceive the above sharp and clear rules for the logical construction of a proposition, then we might say, "these rules must occur in reality"; and, if we realize, despite understanding the relevant propositions, that we don't see how the rules so occur, we might also say, "that isn't how it is." Such contradictory statements characterize a certain disquiet, which disquiet in turn characterizes philosophical problems.[54]

Persistence, or seeming intractability, which also characterizes them, can also arise from an apparent perception. Begin with *PI* §113, in which, being appeared to as above, one may ineffectually repeat to oneself: "But *this* is how it is…" To illustrate, expanding upon Wittgenstein's example, I may think I perceive, although they are out of focus, the sharp and clear logical rules lying in particular everyday definite descriptions; and, I may "feel as though, if only I could fix my gaze *absolutely sharply* on [these] fact[s] and get [them] into focus, I could not but grasp the essence of the matter"; then, I might—again and again—try to capture this essence, i.e., the rules, and try to do so by, e.g., writing this formula: $\exists x(F(x) \wedge \forall y(F(y) \rightarrow x = y) \wedge G(x))$. Next, in *PI* §114, picking up on the above glasses simile, he describes such ineffectual repetition as a misunderstanding: "One thinks that one is tracing nature over and over again, and one is merely tracing round the frame through which we look at it." For example, writing the above formula, we may think we are predicating of certain definite descriptions the sharp and clear logical rules lying within them, but we are, instead, inadvertently only repeating to ourselves features of a certain logical ideal, i.e., one we use to represent such descriptions. Finally, in *PI* §115, he connects this repeated ideal-tracing to a problem's persistence. First, he says that a "*picture* held us captive." For example, given that the above formula pictures the propositions that definite descriptions express, if we repeatedly trace this picture thinking we're capturing the propositions' essence, and if we thereby entangle ourselves in disquieting contradiction, then the picture holds us captive. He goes on, in light of the above can't-step-outside metaphor, to say that language gives rise to such captivity. That is, "we couldn't get outside" the pictures because they "lay in our language," which "seemed only to repeat [them] to us inexorably."

Extrapolation To extrapolate, I'll use each of the above three cases in turn.

I'll use the first to outline the overall approach. It consists in trying to give two sorts of account. One is of how certain philosophical problems about thought experiments arise from language. An account of this sort might be of how the problems arise from explanations required by ideas that themselves arise from pictures of language. The other sort is of how, by clarifying that language, we can solve the problems. Such an account might be of how we can come to see that the ideas so arise, thereby freeing ourselves of those explanatory requirements which lead to the problems.

54. Cf. *PI* §111 & §123.

To illustrate, consider a form the overall approach could take. The problem: How could we possibly learn anything about the world from a *thought* experiment? The problem genesis account might begin with an important picture of terms like "thought experiment." For example, it may begin with a picture of each as an adjective like "thought" modifying a noun like "experiment," or else as an adverb modifying a verb—a picture understood to explain, in terms of these grammatical forms, others, such as "thought experimenter," "thought experimental," "So-and-so's such-and-such thought experiment," and so on.[55] This account may then describe how the picture gives rise to a certain idea, e.g., that a thought experiment must be a kind of experiment that one carries out in thought. Then it may describe how this idea comes to require an explanation, e.g., of how, without observation, or any other source of new information about the world, certain historical thought experiments could possibly have confirmed any empirical theory. Finally, it may point out how, such an explanation proving difficult, the requirement produces the problem. Now turn to the other account, the problem solving one. It may begin with a way of clarifying those terms, e.g., comparing language games involving them, to bring out a picture we use for them and how it gives rise to the above ideas. It may end by pointing out how, with those terms clarified, we can clear problems away, e.g., how, once we understand the terms, the ideas they give rise to no longer lead us, via explanatory requirements, to the above problem. Thereby, one might "solve" the problem—i.e., resist both the idea about thought experiments and, with it, any need to explain how we could possibly learn anything about the world from an experiment performed in thought. To be clear, here, one does not resist this need to explain because one denies that we could so learn. Rather, one denies that we need so much as raise the question.

To develop the problem genesis account, I'll extrapolate from the second case. We may add, to the rise of problems from language, a way of misunderstanding that language. Specifically, we may add that certain forms of expression that we use in connection with thought experiments prevent our seeing what's in plain view about them, thereby misleading us. Consider, for example, the expression "What is a thought experiment?" When we take the question to ask for an analysis—i.e., for us to dig out a hidden essence and not to survey, or get an overview, of its in-plain-view function and structure—then, if we already picture the term as an adjective modifying a noun, or adverb a verb, we're inclined to think thought experiments have an essence that consists, at least, in being a kind of experiment performed in thought. This, together with their importance, might mislead us, that is, lead us to think that they do something extraordinary, even unique. Take, for instance, Galileo's historically important falling bodies thought experiment.[56] It seems, in one fell swoop, to have destroyed Aristotle's theory of free fall and to have established our own modern one; and, this importance, together with the idea that thought experiments must be experiments performed in thought, may well incline us to think that, extraordinarily, Galileo experimented *wholly in thought* to so destroy one empirical theory and establish another. In turn, we might be inclined to think that this extraordinary achievement outstrips, in "justificatory force," any argument from old empirical data and, also, that therefore not all thought experiments in the sciences reduce to such arguments.[57] To be sure, in this example and those below, I'm outlining possibilities to illustrate the extrapolated approach—not trying, as in later chapters, to defend its fruits.

Finally, to further develop the account of problem genesis, and especially of misunderstanding language, I'll extrapolate from the third case. Let us take the above remark that reaps as our starting point.

Sometimes we predicate of thought experiments what lies in our mode of representing them. We do so, for example, when we say that, while performing them, we observe our imaginings, mentally manipulate variables, and test theories—insofar as these lie in our experiment-in-thought ideal. Alternately, prominently, John

55. Cf. Gendler, "Thought Experiments Rethought—and Reperceived," 1154-5.
56. Cf. Galilei, *Two New Sciences: Including Centers of Gravity and Force of Percussion*, 65–72.
57. Cf. Gendler, "Galileo and the Indispensability of Scientific Thought Experiment," 410.

Norton arguably does so in his earliest arguments for necessary conditions on being a thought experiment, those in which he takes "thought experiment" to mean at least "whatever is both thought-like (and so warrants the label 'thought') and experiment-like (and so warrants the label 'experiment')."[58] In so predicating, moreover, we might, dazzled by such an ideal, become confused. For instance, we're sometimes stirred picturing thought experiments as experiments carried out in the laboratory of the mind.[59] This may distract us from the in-plain-view function and structure of terms like "thought experiment," which in turn may lead us, confused, to deny, as we sometimes do, that hypothetical reasoning alone can count as thought experiment.

Also, we are sometimes impressed by the possibility of comparison between a means of representation for thought experiments and cases of them—e.g., impressed how much a given case is like an experiment in thought. Do we sometimes take such impressive comparability to be a perception of a highly general state of affairs? I may do so when—impressed by how well our experiment-in-thought ideal matches my experience with thought experiments—I take imaginings of mine to lie in their nature quite generally, as observation in that of experiment. To illustrate, we can, to some extent, read James R. Brown as doing so in the following passage:

> I have made being 'visualizable' or 'pictureable' a hallmark of any thought experiment. Perhaps 'sensory' would be a more accurate term. After all, there is no reason why a thought experiment couldn't be about imagined sounds, tastes, or smells. What is important is that it be experienceable in some way or other.[60]

Alternately, so impressed, I might take my experience of changing my imaginings to lie in their nature quite generally, as manipulating variables lies in experiment.[61] That is, it may sometimes appear to us that we perceive in reality what is, in effect, an ideal or, specifically, a certain possible comparison thereof. We may then, making a move like the above central one, think that the ideal "must" occur in reality. For example, we might then claim that thought experiments "must" have in them imaginings, or manipulations thereof, or else induction on them to test a theory, and so on. Two points. First, we may make the claim without seeing how the ideal occurs in reality, e.g., how we "observe imaginings," i.e., "introspect them," or "mentally manipulate them." We may make it without so seeing how, moreover, because, ideal in mind, we think we understand terms like "thought experiments" and so know what goes on in actual cases, however it might in fact happen. Second, we may make this move from apparent perception to necessity claim not grasping what that "must" means. That is, we may forget that we need not use a certain ideal, or frame, for terms like "thought experiment." After all, as we'll see in the final chapter, we sometimes alternately represent them by—instead of experiments in thought—arguments, examples, or narrative fictions, among other things; and, doing so, we might say of them that they "must" have conclusions, have the power to exemplify, be narratives, and so on, respectively.

Turn now, from misunderstanding these ideals, to how resultant appearances may yield disquiet and persistence, two characteristics of philosophical problems. First, certain similes are absorbed into our language, e.g., comparisons to experiments in thought that are suggested by the term "thought experiment." These, as we saw, may lead us into thinking we perceive highly general features of thought experiments, such as observing imaginings, mentally manipulating them, and thereby testing a theory. Such appearances yield disquiet if, e.g., we notice that we do not see how we observe imaginings, and so on. To capture this disquiet, we might say, if we think we perceive these features, that, generally-speaking, they "must" really occur in thought experiments; and, if we realize, despite understanding many thought experiments, that we don't see

58. Norton, "Thought Experiments in Einstein's Work," 129-130.
59. Cf. Brown, *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, 1.
60. Brown, 17.
61. Cf. Brown, 17–18.

how they could occur in reality, we may feel compelled to contradict ourselves, i.e., to say they aren't there, or at least that they need not be. Second, persistence, or seeming intractability, may also arise from such an apparent perception and, ultimately, from language. To illustrate, first, we may ineffectually repeat to ourselves that *this* is how it is, e.g., that *this* is a *thought* experiment. We may say so when we feel as though if only we could get this kind of experimenting, which we think we perceive, into focus, then we'd grasp the essence of the matter. Also, to so repeat this may be to keep misunderstanding an ideal in our language. That is, doing so, we may think we are tracing the nature of thought experimenting while only tracing around the frame, or ideal, by which we represent it. Finally, so repeating and misunderstanding, we may be caught in a disquieting contradiction arising from language. We may insofar as an ideal, or a picture, arises inexorably from language and we forget that it can be put aside, like glasses. We might forget it, as we'll see below, because we overlook such simple and familiar pictures, e.g., fail to notice that we need not think of "thought experiment" as adjective and noun, or adverb and verb.[62]

In sum, I've been extrapolating, from three cases, an approach to thought experiments. It consists, so far, in giving two accounts—a problem genesis one and a problem solving one—along the lines of those outlined here. The two form a single approach insofar as the genesis one, by clarifying, helps the solving one.

1.2.2 The Novelty of the Approach

Now, it may be objected, because my approach wields clarification alone, it cannot possibly solve significant problems. After all, at the heart of the literature lie problems like this: How could anyone possibly learn anything about the world carrying out an experiment in thought? And, at best, clarification can help us to better understand such problems, but not to solve them. For thought experimental phenomena call out for explanation, and nothing but an explanation will satisfy us.

But we need not explain to solve. In particular, some problems might have their sources in confusions, ones that make it seem as though there are phenomena that call out for explanation, and clarification might clear up the confusions, thereby clearing away the call for explanation, and so solve the problem—without explaining any phenomena. To be clear, this explaining is not "explaining away," whereas the clarification is. Now, such an approach I have been extrapolating from Wittgenstein's writings. I want now to separate it from two groups of other approaches in the literature.

It will differ from one group, which I'll frame in familiar terms, by not giving such explanation. At the origins of the recent thought experiments literature, John Norton motivated his empiricist position arguing against the platonic one of James R. Brown.[63] Since then, we've often located positions between the two, e.g., as Letitia Meynell does:

> There is also a third approach that is more popular than the other two. Tamar Gendler… Nancy Nersessian… Nenad Miščević… and others have defended a family of views that can be thought of as mental modeling accounts of [thought experiments]. Whether understood as visualized states of affairs… or conceptual schemas… they hold that [thought experiments] are mental kinds—models that we manipulate in our imaginations so as to garner insight or persuade.[64]

Now, the approach I extrapolate below is neither this third approach, nor the first or the second, nor some combination of them. It will differ from them insofar as, following it, first, we should not try to explain how thought experiments yield empirical knowledge or understanding. Rather, second, we should clarify language aiming, among other things, to clear away any need to so explain.

62. Cf. *PI* §129.
63. Norton, "Thought Experiments in Einstein's Work," 129.
64. Meynell, "Imagination and Insight: A New Account of the Content of Thought Experiments," 4150.

The extrapolated approach will, then, resemble logical analysis and Carnap-style explication. So it will resemble approaches in the second group, e.g., Timothy Williamson's and Sören Häggqvist's. But it will differ from these as well, insofar as it has us eschew both "digging for essences" in the phenomena and interfering with actual usage.

These differences, between it and those in the first and second groups, I aim to establish in the remaining two sections, respectively. If successful, I'll have shown the extrapolated approach to be novel relative to certain important positions in the literature.

By the way, this approach isn't the only Wittgenstein-inspired one. I won't examine in detail how mine differs from any others, but I will here point out a difference in focus. That is, mine primarily concerns, in this chapter, extrapolating from Wittgenstein's writings and, in later ones, the nature of thought experiments and how they relate to works of literary fiction. By contrast, Cora Diamond, who makes explicitly Wittgensteinian points about the use of thought experiments, focuses instead on ethics, e.g., argues that to insist, like Roy Sorensen, on stipulation's absolute power is to lose moral relevance.[65] Similarly, Richard Gale, in the recent literature founding Horowitz and Massey collection, deploys Wittgensteinian concepts but focuses instead on using them, somewhat like Kathleen Wilkes' in her well-known book,[66] to identify a class of "pernicious" thought experiments and, unlike her, to repurpose them; for instance, he argues that bizarre science-fiction ones about personal identity fail, since they lack our norm-governed identifying practice, but that we can use them to illuminate such norms.[67]

1.2.2.1 Problem Solving without Theoretical Posits

Again, the approach that I'll shortly extrapolate differs from those in the above first group. These differences are among those that, for Wittgenstein, separate philosophical and scientific investigation. Whereas, for example, a first group approach might have us make theoretical posits to explain phenomena, as in the natural sciences, my approach has us avoid doing so. To establish these differences, and so my approach's novelty, I interpret a key remark, *PI* §109, in light of various others, and then both extrapolate and make relevant comparisons.

Interpretation At *PI* §109, Wittgenstein contrasts his considerations with scientific ones, which his "must not be."[68] One difference is that they're not for theorizing, only describing.[69] Specifically, his considerations cannot aim to "advance any kind of theory," cannot contain "anything hypothetical," and must describe in place of explaining. Another, connected difference is that the considerations, and particularly the descriptions in them, are for solving philosophical instead of empirical problems. This distinction he draws in terms of problem-solving means in two steps. First, they are solved through a certain "insight into the workings of our language," specifically, through recognizing the workings "*despite* an urge to misunderstand them." By contrast, we would not use such a means to answer an empirical question, such as, "What is the specific gravity of hydrogen?"[70] Rather, we might try to make a new discovery, e.g., by measuring the density of a certain hydrogen sample. Second, they are solved through assembling "what we have long been familiar

65. Diamond, "What If X Isn't the Number of Sheep? Wittgenstein and Thought-Experiments in Ethics," 248–249.
66. Wilkes, *Real people: Personal Identity Without Thought Experiments*, 1–48.
67. Gale, "On Some Pernicious Thought-Experiments," 301.
68. Cf. *TLP* §4.111.
69. Cf. *TLP* §4.112.
70. Cf. *PI* §89.

with."[71] Finally, he characterizes philosophy as trying to dispel such problems: "Philosophy is a struggle against the bewitchment of our understanding by the resources of our language."[72] For example, above, to solve philosophical problems about empty names, the nature of propositions, or the sublimity of logic, we struggle to clarify the meaning of empty names, the structure and function of propositions, or a resource of our logical language, i.e., its ideals. This struggle toward clarity is against our inclination—one which arises from pictures of names or of questions or of logical rules—to misunderstand empty names or the nature of propositions or logic's ideals, which misunderstanding leads us into philosophical problems. If we prevail, we do so by means of accounts that provide linguistic insights, ones arrived at by assembling what's long been familiar—e.g., about the workings of names, propositions, or ideals—not by making a new discovery. These accounts, or considerations, moreover, may be composed of descriptions, e.g., ones of long familiar but now overlooked logical ideals, or of the uses and make up of propositions, or of the workings of names without bearers. They may not, however, explain, advance a theory, or make a hypothesis/theoretical posit, as in the sciences—e.g., posit Lewisian possible worlds to explain empty names, or posit abstract logical laws to explain the extraordinary behaviour of propositions, or again posit such laws to explain why logic is sublime.

Consider four clarifications. They concern, in order, how resisting bewitchment bears on theories, how insight differs from discovery, why explanation is prohibited, and what it is to assemble what's long been familiar.

First, the struggle against bewitchment isn't to replace one theory with another, yet it still bears on theories, since it sometimes aims to remove explanatory requirements. For example, the above problem solving account for empty names, if successful, removes any need to explain the nature of their bearers, thereby rendering otiose theories of that nature. To better anchor this in Wittgenstein's writings, first, consider an assertion he criticizes in *PI* §110: "Language (or thinking) is something unique." This assertion, he says, "proves to be a superstition (not a mistake!), itself produced by grammatical illusions." To illustrate, recall the above claim that the strict and clear rules of logic "must" be real (a superstition), which claim arises from our appearing to see those rules in reality (a grammatical illusion), and which appearance depends on tracing a linguistic ideal thinking we're tracing what's real (a grammatical source of illusion). To be sure, it wasn't said that the claim is mistaken, only that its "must" is misunderstood. Now, this struggle against bewitchment may bear on theory, since there's no need to explain with a theory why logic must be so. Now, to better support this reading, it hangs together with the following one. In reply to David Pears' mistaken reading of Wittgenstein, John McDowell approvingly cites Cora Diamond:

> But to attribute [as Pears does] a thought on these lines [i.e., that a certain notion is false] to Wittgenstein is to miss the character of his objection to the idea of the occult mechanism. To echo Cora Diamond, it is to read his "criticism of… mythology or fantasy… as if it were rejection of the mythology as a *false* notion of how things are."[73] If we read Wittgenstein like this, it will seem that the supposedly rejected false notion needs to be replaced with a true one… [But Wittgenstein] objects only if we fall into mythology, and picture that contemporary mental equipment as a configuration in the occult medium of the mind.[74]

That is, to so read Wittgenstein is to take his superstition criticism as if it were attributing a mistake to be corrected by replacing the false notion with a true one. To see how this hangs together with my reading, cast my example in these terms, as follows. Wittgenstein doesn't deny that the idea of sharp and clear rules of logic

71. The third edition has "reporting new experience" in place of "coming up with new discoveries" and has "arranging what we have always known" in place of "assembling what we have long been familiar with."

72. The third edition has "battle" in place of "struggle" and "intelligence by means of our language" in place of "understanding by the resources of our language."

73. Diamond, *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*, 6.

74. McDowell, "Are Meaning, Understanding, etc., Definite States?," 93–94.

corresponds to what exists in reality. Rather, he criticizes how this logical ideal misleads us such that, falling into mythology, we claim that they do so exist. This criticism of the claim, then, doesn't call for its replacement.

Second, Wittgenstein says we solve philosophical problems with a certain insight into language but without any discovery—but how do insights differ from discoveries? To shed some light, consider a nearby remark, *PI* §118, in which he calls such insight "discovery": "The results of philosophy are the discovery of some piece of plain nonsense and the bumps the understanding has got by running up against the limits of language." [75] The idea here, recast in light of *PI* §117, is that successful philosophical investigations result in two sorts of insight—first, that the words you believed to work well—e.g., the sentence, "I'm using [the expression] with the meaning you're familiar with,"—do not in fact work and, second, that the words led you into persistent problems—e.g., led you to keep treating their meanings superstitiously, as if they were auras, and then to entangle yourself over, say, the nature of meanings.[76] To some extent, these insights count as discoveries as do new observations or measurements that test a theory. For both are results. But the insights, unlike discoveries, are recognitions. Recall that we reach them, against an urge to misunderstand, by arranging what has long been familiar. That is, they don't count as discoveries insofar as we do not reach them learning something new—as we do when making new measurements or observations.

Third, why not use explanation—or else theory, hypothesis, or discovery—to solve philosophical problems? In certain cases, as we saw, because explanation is otiose. Beyond that, perhaps because not using it belongs to philosophy, conceived as follows. Consider *PI* §126: "Philosophy just puts everything before us, and neither explains nor deduces anything." Recall that it arranges what has long been familiar. Still, we want to ask, why does it do the one but not the other? He goes on: "Since everything lies open to view, there is nothing to explain." But why don't we have to explain what's hidden? "For whatever may be hidden is of no importance to us."[77] But why isn't it at all important to us? "The name 'philosophy' might also be given to what is possible *before* all new discoveries and inventions." That is, what's now hidden doesn't belong to what we can call "philosophy." To be sure, what's "hidden" doesn't include certain aspects of what's open to view.[78] These include those we recognize after arranging what's open to view but failed beforehand to notice because of their simplicity and familiarity.

Finally, fourth, what is it for philosophers to "arrange what has long been familiar to us" or to "just put everything before us"? Consider *PI* §127: "The work of the philosopher consists in marshalling recollections for a particular purpose."[79] That is, with a certain aim, philosophers assemble and order what is remembered, which is of course, sometimes, open to view and long familiar but neither a new invention nor a new discovery. But whose recollections, or to whom is it long familiar or in plain view? At least anyone doing philosophy. As he says in *PI* §128: "If someone tried to advance *theses* in philosophy, it would never be possible to debate them, because everyone would agree to them." That is, agreement prevents debating philosophical "theses," since they'd be common ground.

Extrapolation Add to the approach so far extrapolated, especially to its account of problem solving, the following directions. It may not advance theories about thought experiments; nor may it contain anything hypothetical, e.g., a conjecture that a future discovery about thought experiments might confirm; nor may it explain, e.g., make a posit to best explain thought experimental phenomena; nor may it attempt to solve

75. The third edition has "uncovering" in place of "discovery," which may sound less scientific but nevertheless retains the here crucial notion of something behind what's in plain view.
76. Cf. *PI* §120 & §125.
77. The third edition treats what may be hidden as an example of what is hidden, qualifying the claim differently.
78. *PI* §129.
79. The third edition has "assembling" in place of the richer "marshalling," i.e., collecting and arranging.

problems by making discoveries, e.g., by making new measurements or observations of brain or behaviour. Instead of doing so, and of explaining in particular, it should describe, e.g., by means of comparing language games, the function and structure of terms like "thought experiment." These descriptions, since they're ultimately for solving specifically philosophical problems, should aim for insights into the relevant language against an urge to misunderstand it. For example, they may aim for insights into the ideals we use for terms like "thought experiment" against an urge, arising ultimately from those very ideals, to think, superstitiously, that a given feature of them "must" exist in reality. Also, to gain such insight, or clarity, we are to assemble and arrange certain recollections—i.e., what is in plain view and has long been familiar. To illustrate, we may recall explanations of what terms like "thought experiment" mean and order them to bring out long familiar pictures, or ideals, that we use to make sense of those terms. Such clarification, finally, as we'll shortly see, may bear on theories of thought experiments by helping us, e.g., to understand such pictures and so to resist ideas arising from them that call out for explanation.

Let us now contrast this approach with each of the three in the above first group. Doing so will bring out its novelty, which lies partly in its aiming to make neither theoretical posits nor discoveries as in the natural sciences. Afterward, I'll reply to two objections.

To begin, recall James R. Brown's striking metaphor: "Thought experiments are performed in the laboratory of the mind."[80] This "bit of metaphor," for him, lies at a frontier, one on the way to saying just what thought experiments are, and beyond which the going gets tough. Instead of sharply defining the term "thought experiment," he then, given that we recognize its bearers when we see them, gives examples, alongside certain hallmark remarks, e.g., that, as we saw, they're in some way experienceable. Subject delimited, later, he gives a taxonomy, in virtue of which he identifies a special class of thought experiments—i.e., ones, such as Galileo's above, which simultaneously both destroy one theory and confirm another, both from unproblematic thought experimental phenomena and without an established background theory.[81] From this class, he argues abductively for his most distinctive position on thought experiments, namely, a theory that posits sense-perception-like intuition of natural laws, understood to be abstract objects.[82]

Conversely, on my approach, we should not advance any such theory, much less argue for it abductively or replace it with another one. We are instead to offer accounts like the above problem genesis one, describing not hypothetical intuition but long-familiar in-plain-view uses of language, e.g., ideals like the experiment-in-thought one, with which that bit of metaphor dazzles us. In connection with this, we might try to give an account of how, from this grammar-spun metaphor, we learn to use that ideal and how, dazzled and confused, we sometimes think we simply recognize, when we see them, thought experiments or experienceables therein. Also in this connection, we may try to describe how, when it comes to tracing classes of thought experiments such as Brown's special one, we're inclined over and over to trace the ideal, or frame, thinking we're tracing phenomena—the very stuff on which his abductive argument for his theory depends—and thereby fall into superstition. Of course, such attempts fail if we can give no such description, and for now I'm not describing. Rather, I'm merely illustrating such description to contrast approaches.

Second, consider a partial sketch of how John Norton, in one publication,[83] gets to his distinctive position, or perhaps it's how a token empiricist might come to it. To begin, for him, we can think a bit and thereby come to know something only if we come to it by "transforming" what we already know. Well, he's not so sure this principle holds for logical truths, and he's mute on mathematical ones, but in any event that's one principle he

80. Brown, *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, 1.
81. Brown, 35–43.
82. Brown, 98–108.
83. Norton, "Why Thought Experiments do Not Transcend Empiricism," 49.

assumes. Here is another, which he's sure of: to transform what we now know into what we will but don't yet know, well, we *must* do something to get from, or likely get from, *the truths* in what we now know to those in our future knowledge. Now, given these two principles—and that thought experiments are bits of thinking by which we come to know things—we can see that they, thought experiments, are ways to transform what we already know; and, in particular, they're something we do to get, or to likely get, from truths in what we know to others in what we will but don't yet know. Moreover, we can do this transforming with arguments, and, for all he knows, we can't do it with anything else. So, what else could thought experiments be but arguments? Norton, then, settles with a thesis about how we learn from thought experiments, that they are arguments. Not that they appear to be what they are, however. Rather, they are "disguised in a vivid pictorial or narrative form."[84]

Here Norton advances a theory, making as he does an explanatory posit, i.e., that thought experiments, hidden under their pictorial or narrative disguises, are arguments. Conversely, again, on my approach, we should not do so. Rather, we should describe what is in plain view, aiming for insight into misleading language, and so on. To this end, my approach may proceed along the following lines. It may begin by examining Norton's initial claim that, logical truths aside, we gain knowledge by thinking alone "only if" we transform what we know. If we take this claim to be a plausible description of how we often learn, might we, misunderstanding the "only if," be tracing an ideal—perhaps one learned in logic class, of sound argument, i.e., valid inferences from true premises—thinking we're tracing the nature of how humans come to know? When Norton goes on to introduce thought experiments, arguing that they must so transform knowledge, do we also use such an ideal to make sense of what the term "thought experiment" means? Then, when he argues in short that, since only arguments are such transformers, thought experiments are arguments, are we readers, who find this plausible, again tracing an ideal, the one we used to make sense of those transformers and the term "thought experiment" in the first place, all the while thinking we're tracing their nature? This illustrates one possible line of approach. Alternately, about this important, impressive comparison of Norton's between arguments and thought experiments, do we sometimes take it to be a perception of a highly general truth about the nature of thought experiments, i.e., that they're arguments which aim to transform old into new knowledge? In this light, when we call certain thought experiments arguments, do we—dismissing as accidental surface features that which doesn't belong to the ideal, especially pictorial or narrative properties—predicate of them what lies in our ideal, i.e., in our mode of representation? Now, to be sure, both of these lines may fail. They would, e.g., if it's no insight into Norton's initial claim that we're inclined to make sense of "transforming knowledge" by means of our long familiar argumentative ideal.

The first group's other positions, as we saw above, have us appeal to a theory of mental models. For example, consider Nancy Nersessian's hypothesis:

> My hypothesis is that executing a thought experiment is constructing and manipulating a mental model. It is a species of reasoning rooted in the ability to imagine, anticipate, visualize, and re-experience from memory. When a thought experiment is successful, it can provide novel empirical data.[85]

Alternately, here is a straightforward part of Nenad Miščević's theory:

> 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 gets updated. The background conditions are dictated by the thought experimenter's general knowledge about the world.[86]

84. Norton, "Why Thought Experiments do Not Transcend Empiricism," 45.
85. Nersessian, "Thought Experimenting as Mental Modeling: Empiricism Without Logic," 127.
86. Miščević, "Modelling Intuitions and Thought Experiments," 194.

Both philosophers try to explain how someone could successfully perform a thought experiment by appeal to mental models. That is, both advance a theory which posits them to solve a problem. Again, conversely, mine cannot advance a theory to do so. To be sure, they appeal not to imaginings, which are in plain view when we try to recall experiences we call "thought experimenting," but to models which may involve them. Also, such theories might solve empirical problems, but these aren't solved by an insight into temptingly misleading language, by describing what's in plain view, and arranging what's long been familiar. So, they do not solve the problems at which my approach aims.

Finally, consider two objections. The first arises from Roy Sorensen's *Thought Experiments*. There, "[he] let[s] the surface grammar of 'thought experiment' be [his] guide" to "understanding" philosophical and scientific thought experiments—i.e., to help him "to establish true and interesting generalizations about them."[87] This may, for example, underlie his definition, namely: "A *thought experiment* is an experiment (see p. 186 [where he defines "experiment"]) that purports to achieve its aim without the benefit of execution."[88] Later, however, he doesn't merely let grammar guide him; rather, he gives arguments. He does so arguing analogically for the thesis, similar to the definition, that thought experiments are experiments—albeit a limiting case of them—or, hedging this thesis, that we ought to stipulate that we use the term "thought experiment" in this way.[89] A key (dis)analogy in this argument is that, unlike thought experiments, "most ordinary experiments are executed and thereby provide fresh information."[90] Now this argument has others at its back and, of particular importance for us here, one for the claim that the term is not a "systematically misleading expression."[91] Were it so, the (dis)analogies might arise from our being misled by the term as opposed to their arising, e.g., from observing how experiments and thought experiments relate to each other. For instance, I take it, if the expression "thought" in the term leads us to think that a thought experiment must be in thought instead of physically executed, and if the term were systematically misleading, then that (dis)analogy between ordinary experiment and thought experiment is a mere illusion. If so, the (dis)analogy doesn't support the thesis that the one is a limiting case of the other. Now, insofar as this line of argument supports following the term's grammar, it may seem that Sorensen preemptively objects to an approach like mine. But my approach isn't his target, for two reasons. First, I don't claim that any such term is *systematically* misleading, although I do claim that it misleads in certain cases. For example, as we'll see next chapter, it misleads when we, inadvertently tracing an ideal for our use of the term, claim that thought experiments must involve imaginings; and, not being systematically misleading, it doesn't do so when, again tracing an ideal, we explain to students why they shouldn't deny that the case in Thomson's Violinist really occurred. Second, even if the term were systematically truth-tracking, my approach would still get some grip insofar as it sheds light on our luck at being led toward truth by the term's grammar. Finally, regarding his analogical argument for *stipulating* that thought experiments are a limiting case of experiments, I pass it over as an objection, since I will deal with similar but stronger ones below.

The other objection arises from Forster's claim that to give a family resemblance account, as Wittgenstein does, is to *reduce* a concept from some features to others, as a behaviourist does.[92] The objection is that, in this light, the extrapolated approach might not differ from the others, especially Norton's reduction of thought experiments to arguments. My reply is that Forster's interpretation, on which the objection rests, is highly contentious, as he admits, especially in light of Wittgenstein's own remarks. For example, "I want to say here

87. Sorensen, *Thought Experiments*, 3.
88. Sorensen, 205.
89. Sorensen, 228–230.
90. Sorensen, 241.
91. Sorensen, 216–218.
92. Forster, "Wittgenstein on Family Resemblance Concepts," 85–86.

that it can never be our job to reduce anything to anything, or to explain anything. Philosophy really *is* 'purely descriptive'."[93]

1.2.2.2 Problem Solving without Analysis or Explication

The approach I'm extrapolating also differs from those in the above second group, and these differences are among those that, for Wittgenstein, separate his method from certain forms that analysis and explication may at least seem to have. Again, we begin with interpretation.

Interpretation Wittgenstein takes his approach's aims to differ from those analysis may seem to have. To see this, consider an analysis-like method of his, its apparent ends, and how they differ from his actual ones. In *PI* §90, certain linguistic misunderstandings arise, among other things, from certain linguistic analogies, and we can remove some of these misunderstandings by substituting one form of expression for another; and, when this substituting "resembles taking a thing apart," we may call it "analysing" our forms of expression. Now, from *PI* §91, this analysis may seem to aim at uncovering, beneath each of our everyday expressions, a single completely analyzed form. And, from *PI* §92, this aim finds expression when we ask what the "essence" of language is. Alternately, from *PI* §113, as we saw, it finds expression in the feeling that, if only we could get the facts, e.g., in certain bits of language, into focus, we could not but grasp their underlying essence. But this isn't Wittgenstein's aim. From *PI* §91, he does aim to "understand the nature of language" (*das Wesen de Sprache*), that is, "its function, its structure," but this, he says, "already lies open to view" and "becomes surveyable through a process of ordering."

Similarly, Wittgenstein takes his approach's aims to differ from those which it may seem to have in light of his concern with misleading language. To explain, consider a means to the above ordering's end, its false appearance, and why the actual one differs. In *PI* §132, we want to establish an ordering, in known usage, for a purpose; and, also for this purpose, we are continually to render perspicuous distinctions obscured by everyday language; and, this activity's purpose, it may seem, is to reform language. But it isn't. Rather, from *PI* §133, we "do not want to revise or complete the system of rules for the use of our words in unheard of ways"; and, the reason for this is that the purpose, clarity, is "complete clarity"—that is, to completely clear away the relevant problems. But why wouldn't such revising or completing clear away the problems? Recall *PI* §109, in which he characterizes philosophical problem solving as struggling against resources-of-language-caused intellectual bewitchments. To have prevailed, you'd have stopped their rise from language as it is. Perhaps revising or completing language doesn't do so. After all, laying down new rules for the use of words opens up the possibility that, when following the rules, we inadvertently entangle ourselves, succumbing to unenvisioned problems of our own making.[94]

Extrapolation As we saw above, certain misunderstandings about words like "thought experiment" might be brought about, among other things, by analogies between expressions in different regions of our language—e.g., by comparing expressions that involve "thought experiment" to others that involve "experiment," "thought," "argument," or "narrative fiction." To remove such misunderstandings, we may, among other things, break the term "thought experiment" into adjective and noun, or adverb and verb, and so on—i.e., analyze it. But, in so doing, the aim, appearances notwithstanding, wouldn't be to uncover a unique fully analyzed form—as if the term in everyday use were blurry and we need only get it into focus to have a clear

93. Wittgenstein, *The Blue and Brown Books*, 18.
94. Cf. *PI* §125.

view of the matter. Neither is the aim to revise or complete our system of rules for using such terms. For the aim, or ideal, is complete clarity, and changing the rules may give rise to yet unenvisioned confusions. Rather, on the approach as I'm extrapolating it, the aim is to grasp what's already in plain view, e.g., the function and structure of words like "thought experiment."

The approach, then, shouldn't formalize thought experiment language to see it clearly—as it were, in high resolution—like Timothy Williamson in *The Philosophy of Philosophy*. For him, thought experiments are arguments plus the imagination; specifically, they "do constitute arguments, but the imagination plays an irreducible role in warranting the premises."[95] And he wants to "achieve a finer-grained understanding of the structure of the arguments that underlie thought experiments."[96] That is, recast in terms of my approach, he wants to see right into thought experiments and get a clear view of the arguments making up part of their essence. To this end, he considers, as a paradigm, Edmund Gettier's well-known thought experiments, and, in particular, he argues for a certain *formalization* of the argument he supposes it partly to be. Passing over details, here is a sample of such formalization. At its outset he summarizes the analysis of knowledge that these thought experiments try to destroy—i.e., that knowledge is justified true belief. Then he writes it "symbolically": "necessarily, for any subject **x** and proposition **p**, **x** knows **p** if and only if **x** has a justified true belief [JTB] in **p**"; and then he rewrites the preceding this way: "$\Box\forall x\forall p(K(x,p) \equiv JTB(x,p))$"[97] This writing and rewriting, if successful, puts the analysis in the language of quantified modal logic—i.e., formalizes it—and thereby, were it to reach a final analysis, gives us a complete understanding of part of Gettier's thought experiments—i.e., a view of that part at, as it were, high resolution. This differs from our extrapolated approach, on which we aim instead to see what's in plain view, not to see into it or to see it in fine detail. To this end, we may, instead of offering a competing formalization,[98] compare language games aiming to shed light on the term's open-to-view function and structure, e.g., how we take it apart to learn its use and how the unanalyzed whole functions when teaching that use, as we'll see next chapter.

Additionally, on the approach I'm extrapolating, we cannot interfere with the actual use of words such as "thought experiment," e.g., argue for a stipulation as Sören Häggqvist does. As we'll see in detail next chapter, he argues, from diversity, for stipulating a use for "thought experiment"; then he argues for a particular stipulation, which amounts, in light of how he aims to revise a concept of ours to make it more precise, to Carnap-like explication.[99] Instead of such theory-directed explication, the approach I'm extrapolating has us order known uses of such terms, without revising or completing our system for them in unheard of ways. We may, for instance, aim to see whether, so revising or completing, philosophers like Häggqvist introduce yet unheard of ways to entangle ourselves, giving rise to new philosophical problems.

This concludes my argument for the extrapolated approach's novelty, completes the approach I've been extrapolating from Wittgenstein's writings, and ends this chapter. In the next two, this extrapolated approach will be a guiding light.

95. Williamson, *The Philosophy of Philosophy*, 188, n. 7.
96. Williamson, 180.
97. Williamson, 183.
98. Cf. Sorensen, *Thought Experiments*, 132–166, Häggqvist, "A Model for Thought Experiments," Ichikawa and Jarvis, "Thought-Experiment Intuitions and Truth in Fiction."
99. Häggqvist, "A Model for Thought Experiments," 58.

2 On What Thought Experiments Are

This chapter has three sections. In §2.1, I clear away a problem, namely, that we have trouble explaining what we know thought experiments to be. To defend this solution, in §2.2, I argue that our concept of them has a family resemblance character. Finally, in §2.3, I support this argument's central claim—namely, that, to them, imaginings are not common.[1]

2.1 On Our Inability to Explain What We Know Thought Experiments To Be

Consider the question, "What is a thought experiment?" If I know the answer, or only think I do, but cannot answer it, I have a problem—one of the form, "I know but can't explain?!" This problem finds expression, for instance, in the question: How could we possibly have any trouble explaining what we know a thought experiment to be? To solve it, we might try to explain it or, as I do, to explain it away.

Here is the plan: first, in §2.1.1, I'll assuage two worries about trying to solve the problem; then, in §2.1.2, I'll criticize a straightforward problem-solving strategy; and, finally, in §2.1.3, I'll defend the roundabout one I follow.

2.1.1 Two Problem-Solving Worries

The first worry is that trying to solve the problem hardly counts as worthwhile, since no one need encounter it. After all, we can, in principle, read an explanation of what thought experiments are and, thereby, have no trouble explaining what we know them to be. To illustrate, consider two connected examples of such reading.

One is that we might read the explanation Ernst Mach gives in his classic "On Thought Experiments." It begins with the following characterization of experiments:

> Man collects experiences by observing changes in his surroundings. However, the most interesting and instructive changes for him are those that he can influence through his own intervention and deliberate movements… If we observe how a child in the first stages of independence examines the sensitivity of his own limbs, we are driven to conclude that man has an innate tendency towards experiment, and that without much looking about he finds within himself the basic experimental method of variation.[2]

He adds a little later: "Experiments guided by thought [as opposed to instinct] lie at the basis of science and consciously aim at widening experience."[3] Now, an idea here about experiments, namely, that we gain wider experience deliberately finding out what happens under various conditions, then goes into the following characterization of thought experiments:

> Besides physical experiments there are others that are extensively used at a higher intellectual level, namely thought experiments. The planner, the builder of castles in the air, the novelist, the author of social and

1. For a chapter summary, see my preface.
2. Mach, "On Thought Experiments," 134.
3. Mach, 135–6.

> technological utopias is experimenting with thoughts; so, too, is the hardheaded merchant, the serious inventor and the enquirer. All of them imagine conditions, and connect with them their expectations and surmise of certain consequences: they gain a thought experience.[4]

In short, to thought experiment is to experiment in a particular way—namely, to gain experience working out what would happen if. Now, one may argue as follows. Anyone can, in principle, read this explanation and, returning to it when asked what thought experiments are, never have any trouble explaining what one knows them to be. So the problem need not arise and, consequently, solving it is hardly worthwhile.

The other example of reading such an explanation, a contemporary one, has Mach's experiment-based explanation as a historical precedent. It is reading Roy Sorensen's definition, seen above in §1.2.2.1. To do so, we might begin with his definition of "experiment"—reading that it is a kind of procedure that (i) must be for answering or raising a question about a relationship between variables and (ii) must vary some of these variables and track responses, if any, in the others.[5] This done, we might move on to that of "thought experiment"—reading that it is a kind of experiment, as defined, but one that is not executed, and one that nevertheless purports to achieve its aim.[6] Finally, we might expand this last definition—interpreting it to say that a thought experiment is a kind of procedure, one presented as answering or raising a question about the relationship between certain variables—not by actually varying some of them and tracking any response that may occur in the others—but, rather, by doing so in thought. Now, again, one may argue against solving the problem. That is, in short, solving it is hardly worthwhile because it needn't arise, since anyone can, in principle, read Sorensen's definition.

To assuage this worry, notice that the problem regenerates itself when we see that explanations differ in the literature. Even here, for instance, Sorensen doesn't give Mach's similar explanation, since he doesn't, e.g., appeal to widening experience. It regenerates because, once we see such differences, we can ask why we have trouble identifying the correct explanation of what we know thought experiments to be. That is, seeing them, a particular version of the problem re-emerges.

Turning from this first worry, the second is that there simply is no problem to solve, since, in serious study, accounts of what thought experiments are can be safely ignored. That is, a substantial philosophical position on thought experiments is, essentially, a position on how they work—how they yield understanding or justify beliefs of a certain sort in a certain domain—and we need not figure out what they are to take such a position. After all, when one wants to examine their workings, one doesn't get hung up, unable to gather samples. Indeed, at the end of the day, accounts of what they are reduce to those of how they work and, at best, serve only to ornament, organize, introduce, or the like.

To assuage this second worry, notice that identifying samples isn't so easy and that this difficulty bears on evaluating accounts of what they are. For example, many literary fictions aren't arguments, and, since it's unclear whether any of them are thought experiments, it's unclear whether any are counterexamples to theories like John Norton's—which account, as seen in §1.2.2.1, explains how certain of them work by, among other things, reducing them to arguments. The point is that, insofar as accounts of what they are might help to identify samples and so cannot be safely ignored, the problem needn't be nothing.

2.1.2 A Bad Solution Strategy

So far, I've tried to assuage two worries, one about the worth of solving the problem, the other about its very existence in serious study. I'll now motivate my somewhat roundabout solution strategy. To do so, I'll criticize

4. Mach, "On Thought Experiments," 136.
5. Sorensen, *Thought Experiments*, 186.
6. Sorensen, 205.

a straightforward one.

The problem, again, is Why do we have trouble explaining what we know thought experiments to be? To solve it, we might deny that knowledge requires explanation. That is, we might (i) affirm we know, (ii) deny we know only if we can explain, and (iii) affirm we cannot explain. That's the strategy I'll now criticize. Specifically, I'll criticize three passes at following it in light of §1.1.2.

Here is the first. We have trouble recalling what we know thought experiments to be, and so we do not, as it were, have it at hand when trying to explain.[7] A slogan: No-Recall Explained, Problem Solved.

But do we really have trouble recalling what thought experiments are? Don't we recall it when we give examples of them, as we so easily do? And don't we often thereby explain it too?

In light of these doubts, consider a second pass at following the strategy. To set it up, consider an ill-fated attempt to efface these doubts about our inability to explain. In short, examples don't cut it. That is, to explain what they are, it won't do merely to give examples or to identify this as one and that as not one—for, to do so, one must give the correct "real definition." One must, that is, to explain what they are, say something like "a bachelor is an unmarried man" or, in some other way, point out the appropriate set of necessary and sufficient conditions. At best, giving examples only shows that one knows what they are, or else it merely clarifies an explanation thereof, i.e., illuminates a definition.

Here is the second pass. We have trouble explaining what we know them to be because it's hard to recall *their definition*. That is, we affirm we know and cannot explain but deny that we know only if we can explain—on the grounds that we know what they are but are unable to recall *the definition needed to explain it*. A slogan: No Definition Recall, Problem Solved.

But, we may now ask, "Why do we have trouble recalling the definition?" In light of this question, we should doubt that the above answer, "Because it's hard to recall the definition," solves the problem—i.e., accounts for why we have trouble explaining what we know a thought experiment to be. This answer, to be sure, may shed some light on the problem, but that may be because it suggests the question, which is itself a clearer formulation of the problem. That is, the "answer" might merely represent the problem.

To set up for the third and final pass, consider an attempt to assuage this doubt. Knowledge is explicit or implicit. If one has explicit knowledge of what a thought experiment is, one has a formulated definition in mind. If one knows it implicitly, one has instead an *unformulated* definition in mind. Therefore, if one tries but fails to recall the definition, one may nevertheless have it in mind. That is, one trying but failing to recall the definition is one trying but failing to formulate it. Also, one who fails to formulate it may nevertheless have it in mind unformulated. Hence, one trying but failing to recall the definition, i.e., to formulate it, may nevertheless have it in mind, unformulated.

Here is the third and final pass. We have trouble explaining what we know them to be because it's hard to give our implicit knowledge explicit form, i.e., to formulate the unformulated definition we have in mind. A slogan: No Definition Formulated, Problem Solved.

But this revised solution may still only represent the problem. To see this, let us recast the problem using these newly introduced terms, i.e., "unformulated," "implicit" and so on. Here it is: Why do we find it hard to explain what we *implicitly* know thought experiments to be? That is, why the trouble formulating our unformulated definition of a thought experiment? If these aren't mere reformulations of the problem—that is, if the "answer" does in fact solve the problem—then these questions should answer themselves—but clearly they do not. By contrast, the questions should be like "Why can't an unmarried person be married?" or "Why can't you see hidden things?" But they're not. Rather, to solve the problem, one would still have to offer an

7. Cf. Meno's Paradox (Plato, "Plato: Complete Works," 80e–85d).

account of how it is that we have trouble formulating an unformulated definition—e.g., appeal to the need for a long investigation or a special insight.

In light of these difficulties following this straightforward strategy, we've some motivation to follow a roundabout one. This we'll now do.

2.1.3 Following the Solution Strategy I Adopt

The above strategy had us ask why we have the trouble, i.e., of explaining what we know thought experiments to be. My strategy, by contrast, has us ask:

1. Do we really have the trouble?

2. What led us to think we have it?

That is, my strategy, unlike the other, aims to solve the problem—i.e., Why the trouble?—not by explaining it but, instead, by calling it into question, that is, by looking into both whether we have it and how we come to think we have it. To recast the contrast, recall that the other strategy has us (i) affirm we know what they are, (ii) deny we know only if we can explain, and (iii) affirm that we cannot explain. My strategy differs along two lines. First, it neither has us affirm nor deny that we know only if we can explain. Second, more importantly, it has us deny that we cannot explain it, i.e., affirms that we can.

By the way, characterizing my strategy as calling the problem into question makes perspicuous how it differs from the other one—but, characterized another way, the difference is hard to make out, giving rise to an objection we'll see below. To explain, recast my strategy as answering these two questions: Can we easily explain what they are? And, if so, why do we have trouble doing it? This done, we can characterize it, like the other strategy, as explaining a trouble to solve a problem. So characterized, again, the difference is hard to make out, but not impossible. To see it, notice how the troubles differ. In the other strategy, it's that of explaining what thought experiments are, whereas, in mine, it's that of realizing we can already easily explain it.

Moving on, here is how I'll follow the strategy. First—to answer, "Do we really have the trouble?"—I'll look for easy extant explanations of what thought experiments are. If found, second—to answer, "What led us nevertheless to think we have trouble explaining it?"—I'll appeal to misleading forms of expression. As a slogan: Explanation Recognized, Problem Exorcized.

To begin following the strategy, the first of the two answers is that, normally, if asked what thought experiments are, we can easily answer. To do so, we may recall an explanation of what they are, or else some material by which we learned it, and may give examples alongside brief descriptions. Given this first answer, the second is that we—who have no trouble explaining what they are—may nevertheless think that we do have it because of misleading language. In particular, we may nevertheless think we have the trouble because we may, conflating a linguistic model with what it models, think that "explanations of essence are definitions," which idea prevents our seeing those easy explanations lying in plain view. These two answers, once filled in, clear the problem away. That is, for those of us who know what thought experiments are, the question, "Why the trouble explaining what we know thought experiments to be?" no longer disquiets us, since, in light of the first answer, we no longer think the trouble exists, and, in light of the second, we can explain how one may have come to think otherwise.

In what remains of this section, I will try to fill in the first and then the second answer. Before doing so, however, consider a couple objections.

First: "But you explain too little, merely how one *may* nevertheless come to think otherwise, not how we in fact do." This objection would have teeth were the problem this one: "How exactly did we, *Alex, Charlie and Jordan*, who know what a thought experiment is, come to have trouble explaining it?" But it isn't. Describing how particular people went astray, to be sure, would help solve the problem, since it would show that one could go astray in that way; but it isn't necessary, since this possibility can be established in other ways, e.g., as I do, by appeal to a certain general form of misleading language.

Second: "Trouble giving an easy explanation, which you say exists, is trouble explaining—but, you say the latter doesn't exist." This objection fails because this trouble of explaining differs significantly from that of explaining what's easily explained—as I touched on above. "But, if you have trouble explaining what's easily explained, you must have trouble explaining." But this isn't the logical truth it appears to be. By analogy, even if you have trouble drinking a cup of tea with your feet, you need not have trouble drinking tea, since you could easily use a hand. Similarly, even if you have trouble explaining in a particularly difficult way, you need not have trouble explaining, if you can explain otherwise, as we normally do.

Filling in the First Answer

My first answer, again, is that, normally, we can easily explain what thought experiments are. We *normally* can in that, as matter of course, we're able to do so and, if misguided, are not. We *easily* can in that we can when prompted and without much effort or any reference materials.

To further fill in this answer, I'll point out, first, certain recallable explanations that we may give. Second, I'll point out certain rememberable situations from which one can put together an explanation. Both sorts of explanation consist in giving examples and describing.

Recallable Explanations In the literature on thought experiments, many philosophers use examples and connected descriptions to explain what thought experiments are.

Some do so largely by means of examples. James R. Brown, for example, aims, in his first chapter of *The Laboratory of the Mind*, to "delimit our subject matter by simply giving examples."[8] He does so *simply* by giving examples as opposed to giving them not alone but alongside *a sharp definition*—which matters here because he also delimits thought experiments with descriptions, ones which fall short of definitions, saying, e.g., as we saw in §1.2.2.1, that they're experienceable and often involve mental manipulation.

Now, one may worry whether this delimiting largely by examples really counts as explaining. It does, arguably, with qualifications. Think of teaching a child what a cow is by reading a farm picture book and naming each of the animals in it; the child will have learned what a cow is when able, normally, to call only the cows "cows." Similarly, Brown, we may think, explains to his reader *what thought experiments are* largely by giving examples, which example-giving comprises pointing them out and calling them what they are; and, his reader will have learnt what they are once able, normally, to delimit them, that is, distinguish them from other things. But, one may point out, we can take delimiting by example to be independent of an explanation of the nature of what's delimited. Think of teaching a child to get the whole milk out of the fridge instead of the skim by pointing at a red-capped jug of it while saying, "whole milk," nodding, and smiling and but then pointing at a blue-capped jug while saying, "skim milk," shaking your head "no," and frowning; the child will then, hopefully, distinguish whole milk from skim by cap colour, even though cap colour isn't what makes whole milk whole or skim milk skim. Similarly, Brown, one may think, gives descriptions and examples to teach his reader to see that by which one can distinguish thought experiments and so delimit them—and does so even

8. Brown, *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, 1.

though that by which one does it isn't that which makes them what they are—thereby *not* explaining what they are. Now, one may indeed think this, but, arguably, that with which his reader does it, at least in part, makes them what they are. That is, that in his examples to which he draws our attention with descriptions—e.g., visualization, mental manipulation, and so on—isn't like the milk caps but, instead, is meant to and does explain, at least in part, what makes something a thought experiment. This might be put as saying that, at least in some cases, Brown's delimiting largely by examples counts as explanation.

Other philosophers explain in a similar way, with examples and descriptions, without saying that that is what they're doing. Consider two cases.

First, take the opening paragraph to John Norton's "Why Thought Experiments Do Not Transcend Empiricism":

> The essential element in experimentation is the natural world. We learn about the natural world by watching what it does in some contrived circumstance. Just imagining what the world might do if we were to manipulate it in this way or that would seem futile, since it omits this essential element. Yet the literature of science frequently leads us to just such imaginary experiments, conducted purely in the mind, and with considerable apparent profit. These are "thought experiments." We imagine a physicist trapped in a box in remote space, that the box is accelerated by some outside agent, and, from tracing what we imagine the physicist would see in the box, we arrive at one of the fundamental physical principles that Einstein used to construct his general theory of relativity. If this can be taken at face value, thought experiments perform epistemic magic. They allow us to use pure thought to find out about the world. Or at least this is dubious magic for an empiricist who believes that we can only find out about the world from our experience of the world.[9]

Norton here introduces his paper's main topic, the apparent "epistemic magic" of thought experiments. In passing, he explains what they are. To be sure, notice the quotation marks in the line, "These are 'thought experiments'." This line signals an attempt either to inform or perhaps remind a reader what counts as a thought experiment. This attempt includes a description and examples; that is, he, before the line, says that they're "imaginary experiments, conducted purely in the mind," ones in which, to "learn about the natural world," we just imagine "what the world might do if we were to manipulate it in this way or that" and, after the line, he illustrates with a sketch of Einstein's famous elevator thought experiment. That is, he explains what they are by description and example, or so we plausibly read it.

Consider two objections. First, we can call this explanation "a definition and illustrating example" and, consequently, since the example merely illustrates, the definition alone explains; that is, the example explains nothing. But, first, if we can so call it, they explain together. To see this, notice (a) that good illustration, like Norton's, improves explanation and (b) that we could give the reverse argument, namely, that, since the definition merely gives that which the example expresses a standard linguistic form, the example alone explains. That is, the burden of proof lies with thinking they don't explain together. Second, Norton's explanation can't be called "a definition"—insofar as it isn't an appeal to an appropriate set of necessary and sufficient conditions. After all, he doesn't use the copula "are" in "These are 'thought experiments"' like an identity sign; instead, he allows other cases also to be picked out by the term. Second, one may object that Norton isn't even explaining what they are here because he doesn't say that, despite appearances, they're arguments—which idea, as we saw in §1.2.2.1, lies at the heart of his well-known theory. This, however, is like objecting that I haven't explained what water is to a child—after teaching the use of the word by pointing at bath water, tap water, puddles, rain, and so on—because I haven't said that it's one or more molecules each comprised of an oxygen atom bonded to two hydrogen atoms and taught the relevant background chemistry. That we call this saying and teaching "explaining what water is" doesn't preclude, in every case, calling the

9. Norton, "Why Thought Experiments do Not Transcend Empiricism," 44.

pointing and teaching by the same name. To be sure, in special cases, only giving chemical formulae, not pointing and calling, will count as explaining what water is—e.g., in a high school class, on an exam which tests the students' grasp of the notation.

Second, in "Mental Models and Thought Experiments," Nenad Miščević explains:

> A prospective buyer who is otherwise interested in science, might have come across some of the famous thought experiments in physics where the subject is typically invited to "imagine an experimental situation" which is being described, e.g. to "picture yourself" as a freely falling body, or a chain placed over a triangular beam, and then asked to imagine various things happening to these objects, or to "do things" with them—link the chain, rotate an imagined object mentally, etc. Eventually the subject reaches a conclusion and "sees" that the body will fall very slowly, or that the chain will stay still, etc.[10]

This passage's explanation—beginning at "where"—initially separates *saying* from *examples of* what they are, like the Norton passage above, but then, unlike it, the two run together in the second sentence. That is, in the first sentence, Miščević separates what's said and exemplified, doing so with an abbreviated *exempli gratia* and em-dash, but, in the second, he characterizes what, at a thought experiment's conclusion, one "sees" entirely by means of examples and an "et cetera."

But, one may object, this passage doesn't explain, by means of examples and descriptions, what thought experiments are, because it merely explains what goes on in certain famous ones which this "prospective buyer" might have come across. In response, we can see that Miščević aims more broadly, first, from the adverb "typically," in "the subject is typically invited," which would be strange to use here unless thought experiments in general were in question, and, second, from the explanans itself, since it consists in properties arguably too general to concern so specific a class of thought experiments.

These three explanations in the literature—Brown's, Norton's and Miščević's—we may recall them, and, when asked to explain what thought experiments are, we may give them or something like them. Furthermore, giving descriptions and examples in this way—that is, giving them in virtue of recollecting explanations like the three we've gone through—is evidently one normal and easy way we can explain what thought experiments are. In support of this being easy, from reference materials, examples of thought experiments come effortlessly to mind, and, with an example before us, it's fairly easy to recall and describe characteristic features. It's like drawing a horse from memory and then, while looking at your drawing, easily recalling horse features you've been taught are typical. Finally, in support of this sort of explanation being normal, this section's explanations are both often read and typical of many others in the field.

Rememberable Situations Let us leave the case of giving an explanation we recollect and turn to that of giving one by recalling something else. The something else I will consider consists in material by which we have learned what thought experiments are, and, to explain using this material, in short, we give it to someone else. I'll now sketch some of this material and its explanatory use.

I'll sketch it in light of certain initial thoughts about thought experiments, ones which undergraduate philosophy students may well have while still getting a handle on what they are. Also, I'll illustrate these thoughts in a short dialogue, drawn from typical ethics classes, between two students who, well into a typical course on moral philosophy, find themselves using the term "thought experiment" for the first time. This course has, for the last month, been on Utilitarianism, and, naturally, the students have encountered many thought experiments in the following fairly standard way: first, after hearing them summarized, they're asked to evaluate them, the ensuing discussion guided by the professor; second, many of the most memorable ones have purported to destroy this moral theory and have the form *in this case, the theory tells you to ϕ but*

10. Miščević, "Mental Models and Thought Experiments," 215.

intuitively you should not-ϕ; third, many of them have been called "thought experiments," during, for example, an introduction to or a review of class material, but neither has the professor explained the term nor has either student looked it up.

STUDENT A: You don't know about "thought experiments"? Like taking everyone who's a big drain on the healthcare system and secretly euthanizing them, somehow, to make everyone happier, everyone overall, because then there'd be more money to go around?

STUDENT B: Yeah, or there's the electrician being electrocuted, whose suffering keeps a big game's live feed going, a broadcast which delights so many millions of fans that, overall, there's a higher ratio of pleasure to pain than there would be were you to stop that suffering—or there's that "utility monster" one, that individual who derives so much pleasure from any given good relative to everyone else that the greatest overall happiness would be achieved by giving every good to that person instead of anyone else.[11]

A: Right, or there's the possibility of making a world like Hell's populous pagan circle, one in which everyone leads a life of sighs, one barely worth living, but in which there are so many people that the accumulative effect is that it, this world, would contain more happiness than another, say, one like Heaven's first circle, in which there are fewer but happier people—or there's the Utilitarian at a human rights rally who can't explain how anyone could possibly have a right *an absolute right* to anything—oh and there's that experience machine, the one into which you could plug, like into the Matrix, and, but, when you plug in, you get way more pleasure than you would otherwise, like the pleasure of having written a great novel, though you wouldn't in fact be a great author, or even really an author.[12] But what about them?

B: I said, "I don't know about them." They don't really *do* anything. Don't destroy Utilitarianism or whatever. There's always some work around, around saying that a Utilitarian would do the bad thing. I mean, a Utilitarian need not say *yes* to the euthanizing cruise, to letting the electrician be electrocuted, or to giving that monster all the utility—if one, say, rejects "Act-" in favour of "Rule-Utilitarianism"—and, you know, also, because the Utilitarian at a human rights rally who distinguishes between the principle of utility and one's criterion of action can explain that, well, although there aren't any human rights per se, promoting them produces the greatest overall happiness, and ah…

A: Sure, and because you can get around saying that the heavenly world is worse by distinguishing between low and high pleasures and saying that no amount of hellish low pleasure can amount to the heavenly high stuff[13]—and because you can get around saying that you should plug into the experience machine by saying that Utilitarianism is a theory for the real world, not for what's merely possible, and, in the real world, experience doesn't come apart from the way the world is, as it does in the case of the experience machine, and so, for example, you only get the pleasurable experience of being a great author by really being one.[14] I get it. Yeah. You're saying that maybe thought experiments don't work because, well, the ones we've seen, they never really sink the ship.

These students are explaining to each other what thought experiments are. They are doing so by wondering what the term "thought experiments" means and then recalling examples, which they give by sketching

11. Cf. Nozick, *Anarchy, State, and Utopia*, 41.
12. Cf. Nozick, 42–45.
13. Cf. Mill, *Utilitarianism*, 16.
14. Cf. Donner, *The Liberal Self: John Stuart Mill's Moral and Political Philosophy*.

memorable features. That done, one student comes to understand the other's doubt that they ever succeed. In light of this explanation and doubt, we can, at first pass, say that they *understand* what a thought experiment is. Either they already possessed the concept or come to do so during their discussion. But, in light of, as it were, their conception's unfinished contours, we can also say that they do not understand. To see this, notice two such contours. First, take a broad view and notice that neither student distinguishes between thought experiments in different fields, e.g., in science and philosophy or in ethics and metaphysics, and they speak as if no such differences exist, i.e., when expressing skepticism about thought experiments in general and not, e.g., those outside the sciences. Second, instead of a broad view, take a close up one, and notice that, to refer to a thought experiment, neither do the students use proper names, such as "Nozick's Experience Machine," nor do they label characteristic thought experimental features, e.g., say that the "hypothetical scenario" consists in an experience machine, or that the "intuition" we're to have is not to plug in, or that the "outcome" is that more matters to us than how things feel from the inside, and so on; rather, the students recall to each other, in rough and ready terms, some features of various thought experiments that immediately come to mind, e.g., the euthanizing cruise case and the term "utility monster." More generally, neither student distinguishes a thought experiment from the memorable feature described, which leaves unclear where, after this feature, a thought experiment ends, e.g., at the memorable feature, at the case comprising all the features, at a use of the case, at a particular use of the case, at one of these points here and another there, or what exactly. In connection with this, neither distinguishes between their aims, e.g., theory construction vs. destruction, acquiring conceptual vs. empirical knowledge, or challenging a theory vs. disproving it—which gives rise in part to their treating thought experiments as if each were supposed to be a very general mathematical proof, e.g., a disproof of every possible form of Utilitarianism. Now, we may be tempted here, in light of the students' conceptual competence as well as their unfinished concept, to qualify—that is, to say either that the students have a *rudimentary* understanding or that they *nearly* understand. And one may be tempted to argue for saying one or the other. For instance, so tempted, one may argue that the students must understand, if only rudimentarily, because they rightly refer to thought experiments. Furthermore, students in general quickly acquire the concept after only a few examples.[15] Alternately, again so tempted, one may argue that they do not understand, but at best nearly do, because they're explaining and you can't successfully explain to someone what they already understand. But to opt entirely for one option or the other runs afoul of the facts—i.e., that we can say the students understand and can say that they do not. The temptation subsides when we describe these facts as follows: We used unmade distinctions as a standard to judge that they don't understand—judging, as it were, from our more sophisticated, downstream perspective that their concept is unfinished, incomplete—and we used their competence, shown in their example-driven explanation and doubt, as a standard to judge that they do understand. We might call this an intermediate case of understanding—one, as it were, between clear cases of understanding and not understanding—what thought experiments are.

Now, throughout this intermediate case, the students are learning. That is, they are moving between a clear case of not understanding, long before the dialogue's start, toward a clear case of understanding, which might be achieved long after the dialogue's end. Doing so, they make use of learning materials, not classroom pencils and paper, but thought experiments. These learning materials we find at work as the students explain to each other what thought experiments are. Specifically, we find them at work in their assembling and arranging of recollections. To see this, first, notice the simple order they give to their recollections—i.e., one thought experiment recalled by describing a memorable feature, then another, then another, and so on. Second, notice that this order shows some, as it were, contours of their concept. By analogy, if, asked what feathery pets are,

15. I owe thanks for this observation to Jim Brown.

you say, "parrot, canary, budgie, parakeet, and so on," then you may well be surprised or skeptical when told that a crow is a feathery pet; similarly, the students, after assembling and arranging, may well, if they continue on as typical philosophy students, be surprised or skeptical when a professor calls Rawls' veil of ignorance a thought experiment—since its aim, unlike that common to the ordered examples, is to establish something, i.e., principles of justice, and not, on the face of it at least, to destroy a theory.[16] In sum, these thought experiments, which the students assemble and arrange to explain their concept, are materials for learning what thought experiments are.

Such learning materials we can, much later, give away to explain what thought experiments are. That is, we can recall examples we ourselves used, like the students, to work out what they are; and, giving those examples, i.e., assembling and arranging them, we can, again like the students, show the contours of our concept, i.e., explain them. In brief, we can explain what we know with that by which we came to know it. By analogy, we can give someone our necessary-and-sufficient-conditions-concept-building materials when we explain such conditions as we had them explained to us. We can do so by appeal to a definition like "bachelors are unmarried men"—by, that is, using the definition not to explain, or not solely to explain, but as a model on which someone can cut their teeth.

Finally, explaining in this way is both easy and normal. First, the examples need no reference materials and take little effort either to arrange, since the ordering may be simple, or to assemble, since the descriptions by which we give them tend be memorable—e.g., concern something that is shocking or weird or simply concrete instead of boring or common or abstract. Second, as with explaining what something is in general, we commonly do give examples we encountered when learning what thought experiments are to explain what they are. To be sure, we may, as above, add a brief description to the examples we give, and this doesn't make the resulting explanation either unusual or difficult to give. With the examples in front of us, the description isn't hard; and, we often feel we have to give such a description to be understood, that is, to have our hearer's attention directed toward the pattern in the ordered examples which we think is important.

***

Consider four objections to this illustration of how we may, normally and easily, with our own learning materials, explain what we know thought experiments to be.

First, at best, by appealing to such building materials, we can only give a student-level understanding of what thought experiments are—not the full-fledged, sophisticated one at which we might aim. In reply, what if, after appealing to your initial learning materials, you kept on explaining, appealing to later materials, that is, giving revised orderings of examples or descriptions? That is, in principle, you may, by giving all your learning materials, or certain relevant ones, get across your very conception.

But then, second, do the explanations really explain what thought experiments are? That is, are they successful apart from getting across what one understands them to be? If not, it may be objected, I ignore the problem. That is, one may object: Sure, normally, we can easily explain our conception of thought experiments to someone else—but that doesn't assuage my worry about having trouble explaining what they are beyond my conception of them. This objection overlooks that we're assuming ourselves to know what thought experiments are. So, if the explanation gets across our concept—i.e., what we know—then it will be successful in this desired way.

But then, third, one may object that explanations by ordered examples, even ones supplemented by a brief description, will likely be vague or otherwise imprecise. By analogy, merely giving assorted water

16. Cf. Rawls, *A Theory of Justice*, 136–142.

samples—even ones supplemented with the brief description, "the matter is the same in each"—likely won't tell anyone that water is H_2O, since the samples will likely have more in common than simply being H_2O, e.g., will all likely have in them water that's in chemical equilibrium with OH^- and H_3O^+. So, even if we do know what they are, we likely won't, by giving such explanations, express our knowledge and so explain what they are. This assumes, however, that our concept of a thought experiment, or our knowledge of it, isn't itself "vague or otherwise imprecise"—and this is true insofar as it means that it differs from concepts explained, e.g., with a sharp definition—but it's false insofar as it means that it doesn't have a family resemblance character, as explained in §1.1.2. So, even if it's certain that the concept we express with our example-driven explanations is "vague or otherwise imprecise," the explanation may nevertheless perfectly express our knowledge of what thought experiments are.

But, finally, fourth, does our concept of a thought experiment have such a character? I argue that it does, next section, thereby defending the solution strategy I'm now following.

Filling in the Second Answer

The second answer, recall, explains why we overlook the first. That is, it accounts for a misunderstanding of our language, one that leads us to think that we cannot, normally and easily, explain what thought experiments are. To elaborate, first, I'll explain its form, in line with §1.2.1, and then I'll fill in its content.

The Second Answer's Form If we call some expression an "explanation of what something is," we might explain why we can do so. And we might explain it by appeal to similarities or differences between it and a definition, e.g., "A bachelor is an unmarried man." If we do so, we use a definition as a model.

We can confuse this use with an illustrating one. To see how, consider the following explanation. We can call such-and-such expression an "explanation of what something is" because, like the definition "a bachelor is an unmarried man," it specifies necessary and sufficient conditions. Suppose this explanation has the form: We can call it so because it has this property like this model. Now, we may confuse this form with the following similar one: We can call it so because it has this property, which this definition illustrates.

Confusing this use may make various explanations of what something is look essentially the same. In particular, if, as above, we confuse using a definition as a model with using it merely to illustrate, the property illustrated may appear to be not *a* but *the* one which makes expressions such explanations. Suppose we confuse a model use of the bachelor definition, one meant to explain why we call a given expression "an explanation of what something is," with an illustrating use, one which appeals to the property of being necessary and sufficient conditions, which the definition illustrates. We may then overlook alternative uses of the model, e.g., uses of it to explain that the expression so counts because, like the model definition, it simply says what something is. We may also even overlook the use of other models, e.g., uses of it to explain that the expression so counts because—like the model explanation "pets are dogs, cats, goldfish, etc."—it helps someone give further examples of the same kind. Overlooking these other uses, that in virtue of which we think an expression is an explanation of what something is may appear to be unique—one, not many. For instance, if we think the property of specifying a set of necessary and sufficient conditions is that in virtue of which we call an expression an explanation of what something is—and, if, moreover, we overlook other explanations of our calling it so, like those just noted—then, this property will appear to us to be *the* one which explains it. In short, the model so misunderstood ends up being like a pair of glasses through which we look at explanations of what something is but which we don't think to remove.

Now, absently wearing the glasses, as it were, we may trace our model trying to trace the nature modelled.

That is, first, if, confusing the use of a definition as a model with another use, explanations of what something is all look essentially the same to us—and, second, if we want to describe such explanations—we may then describe features of the model we're using to represent them instead of them themselves. For example—tracing model rather than nature modelled—we might say, "an explanation of what something is *must* capture the properties necessary and sufficient for something to count as what's explained," and, if we, as we sometimes do, take capturing such properties to be definition, we might add, "explanation of what something is *just is* definition." Similarly, we may reason as follows: "an enumerative definition *must*, by enumerating instances, specify necessary and sufficient conditions, since it's an explanation of what something is"; or, "an enumerative definition *must not* be an explanation of what something is, or not be an acceptable one, because it doesn't specify necessary and sufficient conditions." This, to be clear, isn't to say that we cannot use these descriptions without having confused a model with what's modelled. These uses of "just is" and "must" mark, as it were, that we're describing model instead of nature modelled, and we might indeed know this and mean to do so, e.g., say we're doing so in an explanation of what we mean by "expression of what something is."

Finally, if we're asked what something is, but, as above, we confuse, with illustration, our use of a definition as a model to call an expression "an explanation of what something is," then we may, tracing model instead of nature modelled, count as an answer only an expression which is, e.g., a definition given in terms of necessary and sufficient conditions. If, then, taking ourselves to know the answer, we try to give or find one, and if, among those expressions which we count as answers, none satisfies us, we will, upon reflection, think we have trouble explaining what we know the thing to be—i.e., will have that sort of I-know-but-cannot-explain problem. Crucially, we may have such a problem even if we could easily explain what the thing is by means of what we, confused, do not count as an answer.

The Second Answer's Content I'll now fill in the form's content and sum up.

Consider the question, "What are thought experiments?" If we know the answer, and take ourselves to know it, we may nevertheless think, first, that we have trouble giving an acceptable answer and, second, that, consequently, we have a problem—namely: Why do we have trouble explaining what we know them to be? To solve *part of* this problem, I argued that we do not in fact have trouble answering, since, normally, we can easily explain what they are—e.g., by means of examples and brief descriptions, which we can acquire by recalling an explanation or certain learning materials.

Now, to solve the rest of the problem, I'm to explain how, even though, normally, we can easily explain it, we may nevertheless think we have trouble doing so. Here is the explanation. We may, using sharp definitions as models in calling expressions "explanations of what something is," confuse model with example; and, thereby, we may see a feature of a certain definition, such as being a set of necessary and sufficient conditions, as essential to explanations of what something is; then, inadvertently tracing our model thinking we're tracing nature, we may be under the impression that explanations of this kind must have that feature—and, if asked what thought experiments are, will take to be an answer only that which consists in necessary and sufficient conditions. Finally, unable to give such an answer that satisfies us, we may think we have trouble explaining what we know them to be—even if, normally and, in particular, when not so confused—we can easily explain it.

2.2 'Thought Experiments' Form a Family

Wittgenstein once wrote: "I want to give an account of the motley of mathematics." [17] I want to do something similar. In particular, I want now to give an account of our unreflective concept of a thought experiment on

17. Wittgenstein, *Remarks on the Foundations of Mathematics*, III.48.

which it has a "family resemblance character."

Giving the account adds to my overall project in three ways. First, doing so, I defend the last section's solution and, in particular, its claim that, normally, we can, with example and description, easily explain what we know thought experiments to be. After all, as we saw, if the concept has such a character, then, arguably, even if, so explained, it's "vague or otherwise imprecise," we may nevertheless perfectly express what we know. Also, second and third, giving the account extends last chapter's approach, especially the line in §1.1.2, and lays groundwork for the next.

This account has three parts. The first, §2.2.1, argues that the concept satisfies one condition on having a family resemblance character, namely, not being sharply definable. In light of this argument, the second part, §2.2.2, contrasts my position on when to "define" the concept with others in the literature. This sets up the third part, §2.2.3, in which I "define," or survey, the concept, aiming to shed light on how it satisfies the other condition—i.e., being explained by family resemblances.

2.2.1 Satisfying the No-Sharp-Definition Condition

A slogan for this first argument: Thought experiments are essence-free, for we see no commonality. And here is a summary. Thought experiments may have properties in common, but, when we examine their properties and explanations of what they are, we see no set of common properties that uniquely explain why we unreflectively call something "a thought experiment." That is, we see that they have no essence and find that no sharp—i.e., no non-vague and non-disjunctive—definition captures our unreflective concept of them. Their concept then, in light of §1.1.2, satisfies the no-sharp-definition condition on having a family resemblance character.

Here is the plan. First, to introduce matters, we'll look for relevant explanatory commonalities in two definitions. Not finding any, second, we'll examine an argument due to Tamar Gendler and Sören Häggqvist—one from vagueness against thought experiment being sharply definable. Finally, third, we'll improve this argument—basing it on seeing *not* common an "obviously essential property," i.e., involving imaginings.

Two Suggestive Definitions

Consider a definition we may well turn to wanting to know exactly what "thought experiment" means:

> OED **thought experiment** n. an experiment carried out only in one's imagination; a mental assessment of the implications of a hypothesis; = GEDANKENEXPERIMENT n.[18]

The definiens comprises three clauses, divided by semi-colons. Understanding any given clause will be enough to satisfy some dictionary users. Others will require the clauses to cohere, e.g., that each one captures, in different but compatible ways, the essence of thought experiments. Yet others will require that both clauses be understood but not that they cohere, e.g., will allow them to have incompatible meanings, although not so different that they require separate dictionary entries. Other possibilities exist, but these suffice to make my point. The definition explains our use of the term—i.e., specifies criteria for calling something "a thought experiment"—and in so doing it refers to properties thought experiments have, but it need not be read as making these properties common. One might object that the definition user who requires inter-clause coherence *must* disagree, but this isn't relevant, since the user needn't be correct. The upshot: we see, in the OED definition, no explanatory common property, only ones that may or may not be so interpreted.

18. *Thought Experiment*.

Now consider another explanation, Ronald Laymon's. We find it in the Horowitz and Massey collection, one foundational to much of the recent thought experiments literature. Here is the explanation:

> A *thought experiment* is an ordered pair $< \Phi, \vartheta >$ where Φ is a set of persons (audience and/or presenter) and ϑ is a set of statements $\{T, P_1, P_2 \ldots P_n , Q\}$ where:
>
> (1) T is a description that is not in fact true (because it is idealized) of any experiments in this world.
>
> (2) Members of Φ believe that $P_1, P_2 \ldots P_n$ are scientific laws or principles.
>
> (3) Members of Φ believe that $\exists\ (Tx)\ \&\ P_1, P_2 \ldots P_n \Rightarrow Q$.[19]

Unlike the OED definition, we read this formal definition uniquely as doing its job in virtue of specifying a common property—i.e., being such-and-such kind of ordered pair. But this is stipulation. Specifically, as Laymon goes on to say, he is "not trying to specify conditions that can be used to specify ordinary scientific usage of the expression 'thought experiment'… [but] to mark off a natural scientific practice that is of scientific importance and of philosophical interest." The point is that we see, in this definition, an appeal to a common property, but it isn't for explaining even ordinary scientific usage, much less an unreflective concept.

In neither definition, then, do we see a common property explaining our unreflective concept of a thought experiment. Let us turn now to the unimproved argument, the one, from diversity, against sharply defining this concept.

Unimproved Argument

Begin with some background. In Kuhn's influential paper, "A Function for Thought Experiments," he remarks, first, that the "category 'thought experiment' is… too broad and too vague" for one instance to stand for all and, second, that many of them differ from the one he examines.[20] That is, the category is like a big ragged-edged box of diverse things.

Perhaps Kuhn's account of the category influenced the contemporary literature. For example, we find such an account in Horowitz and Massey's introduction to their well-known collection, mentioned above.[21] Specifically, of Nicholas Rescher's chapter, they say, first, that, taking a broad view, it just about identifies thought experiment with hypothetical reasoning. That is, as it were, the box is big. Second, they point out an "imperfection" in his taxonomy of thought experimental structure and function, namely, that its cells aren't either mutually exclusive or jointly exhaustive. As it were, the big box has rough edges. Finally, about this charge of imperfection, they assert that the roughness counteracts a seemingly rampant belief in the sameness of thought experiments. That is, the big rough-edged box has diverse contents.

Later, reviewing this collection, Tamar Gendler brings out and develops a connection between such diversity and roughness.[22] The idea, at first pass, is that, since thought experiments differ greatly, they're indefinable. That is, to explain why, in the collection, few philosophers give "definitions" of thought experiments in philosophy—ones like Laymon's, quoted above, and James R. Brown's "Thought experiments are performed in the laboratory of the mind"[23] —she airs the idea that a sharp one cannot be given, or it would be a challenge to give, because of how diverse, or how central to philosophy, the philosophical techniques called "thought experiments" are. As evidence of this diversity, she cites the following views:

19. Laymon, "Thought Experiments of Stevin Mach and Gouy: Thought Experiments as Ideal Limits and as Semantic Domains," 168.
20. Kuhn, "A Function for Thought Experiments," 24.
21. Horowitz and Massey, "Introduction," 2–3.
22. Gendler, "Tools of the Trade: Thought Experiments Examined."
23. Brown, "Thought Experiments: A Platonic Account," 122.

> Nicholas Rescher argues that much Presocratic reasoning can properly be understood as thought experimentation; Peter King contends that medieval treatises on obligatzones (formalized debates or disputes) represent "a developed body of reflection on the method of thought experiment;" Rolf George argues that thought experiments are a defining feature of early modern epistemology; J.N. Mohanty suggests that the Husserlian technique of eidetic variation is basically that of thought experiment; and Gerald Massey argues that thought experiment is contemporary analytic philosophy's main modus operandi, the modern surrogate for meaning analysis.[24]

Later still, Sören Häggqvist, citing Gendler's review, makes the following argument. What has been called "a thought experiment" varies greatly. It ranges, for example, "from mathematical arguments, pre-Socratic reasoning and Husserlian eidetic variation to Harvey's discovery of the circulation of the blood"; therefore, any attempt to give a general characterization of them, as one might give for members of a natural kind, "seems both daunting and misguided."[25] Our box's top, as it were, has let in such variously shaped things that tracing it with a single line appears not just hard but wrongheaded.

This Gendler-Häggqvist argument runs into several difficulties. First, recall that some natural kinds vary greatly. Water does. What we call "water" ranges, for example: in state, from glacial ice to liquid ocean to cirrus cloud; in volume, from a great lake to a rain drop to a single molecule; in purity, from muddy puddle to tap water to what's finely filtered; in use, from oxygenator and hydrator to medium for swimming and floating. Also, we can define it by appeal to a common property, as we've been taught. Why, then, should diversity at all deter our giving a general characterization of thought experiments or at all justify us in thinking doing so is misguided?

Well, perhaps because, here, "diverse" means "lacks commonalities." That is, when wanting to give a general characterization of some things, and so looking for certain common properties, ones which explain what they are—but not seeing any, or sufficiently many, amongst copious differences—we may say, "they're diverse," or, "they've very little in common," or some such. Moreover, this looking for but not seeing significant commonalities counts as evidence that they, the commonalities, do not exist, which evidence should both deter our attempting a general characterization and show it to be misguided—if our evidence suffices.

But does our evidence suffice? That is, do we see no, or so few, explanatory commonalities across Häggqvist's and Gendler's diverse examples? No. For the examples are not all cases of what we call "a thought experiment" but, as Häggqvist puts it, of what has "on some occasion" been so called—and within this wider class we find mistakes or stipulated, technical uses.[26] By hyperbolic analogy, the argument is bad like the following: we cannot define squares sharply because they're so very different; for example, some are equal-sided and square-angled quadrangles, some are rectangles which, by stipulation, I call "squares," and some are dogs.

To illustrate, consider two of their examples: J.N. Mohanty calling Husserlian eidetic variation, and Nicholas Rescher calling some presocratic reasoning, "thought experiment." The two, by the lights of the descriptions by which they're referred to, seem strikingly different from each other as well as paradigmatic thought experiments, but one rests on a stipulation and the other either does as well or may be a mistake. Let us examine each in turn.

Eidetic variation, as Mohanty explains it, is a method for understanding a phenomenon's essence. To follow it, very roughly, first, you vary an imagined instance of a phenomenon until it's another kind of thing, then you repeat the process over and over again, varying the instance in different ways until it's some other kind of thing, and, finally, eventually, you'll grasp what's common to these variations, i.e., grasp the

24. Gendler, "Tools of the Trade: Thought Experiments Examined," 82–3.
25. Häggqvist, "A Model for Thought Experiments," 57–58.
26. Häggqvist, 57.

phenomenon's essence. Following this method, Mohanty thinks, counts as thought experimenting, in part because it's "a process which cannot be reiterated physically," i.e., isn't carried out "instead of, or prior to, actually performing a physical experiment."[27] This reiteration condition is, on Mohanty's final analysis, stipulated; in his Postscript, he argues that, since the term "thought experiment" has no "ordinary extension," and since such physical reiteration isn't philosophically significant, his restriction of the term's extension to what isn't so reiterated is, at least in philosophy, "neither unduly restrictive nor conflicting with ordinary usage."[28] To be sure, the argument has some doubtful premises. For example, some philosophers such as Mach, do appeal to physical realizability to justify using thought experiments,[29] which makes it philosophically significant, and the term has a popular usage which, if not ordinary, isn't quite technical either, in light of the OED definition above. Nevertheless, unjustified stipulation is stipulation.

Similarly, Rescher does call certain forms of presocratic reasoning "thought experiment," but he does so in light of a broad definition for which he doesn't argue. In short, he says, they're attempts to learn by reasoning from a hypothetical.[30] He then applies the definition. For example, he applies it to some reasoning attributed to Thales. Here is the result:

> —To show: the earth floats on water [like a log].
>
> —Assume this to be so, that is, suppose that the earth floats [like a log] on a large body of water.
>
> —Note that this supposition will naturally explain the earth's remaining in its place in nature [and does so at least as well as any available alternative].
>
> —Therefore: we are justified in claiming that the earth floats on water [like a log].[31]

My point is that he counts this presocratic reasoning as "thought experiment" in light of a stipulated definition or else a contentious and possibly mistaken one. Its contentiousness lies at least in its breadth. Andrew Irvine, for example, thinks so: "Surely some instances of hypothetical reasoning, whatever their accompanying intentions, are simply not scientific enough to be deemed thought experiments"—i.e., do not "stand in a privileged relationship both to past empirical observations and to some reasonably well-developed background theory."[32] To be sure, even if Irvine's requirement is much too strong, Rescher's definition and its application hardly escape contentiousness and possible error.

These difficulties with stipulation and possible mistakes may arise, if only in part, from looking at *unusual* things called "thought experiments." To give an improved argument from diversity, let us instead first look at paradigms and a property of theirs, as follows.

Improved Argument

Think of the following sentence as explaining, in part, what thought experiments are: "Well, they involve imaginings." Is a common property appealed to? At a glance, yes, and, indeed, if there is a set of common explanatory properties, this one is a member. But is it really common?

To check, let us begin by examining paradigm cases. Doing so, we find some confirmation, since unsurprisingly such cases usually involve hypotheticals, imaginings, suppositions, and so on. Consider a few examples. First, w.r.t. hypotheticals, there are hypothetical moveables or stones in Galileo's Falling Bodies Thought Experiment: "Then if we had two moveables whose natural speeds were unequal, it is evident that

27. Mohanty, "Method of Imaginative Variation in Phenomenology," 263.
28. Mohanty, 271.
29. Mach, "On Thought Experiments," 136.
30. Rescher, "Thought Experimentation in Presocratic Philosophy," 31.
31. Rescher, 33.
32. Irvine, "Thought Experiments in Scientific Reasoning," 149–150 & Horowitz and Massey, "Introduction," 13.

were we to connect the slower to the faster, the latter would be partly retarded by the slower, and this would be partly speeded up by the faster…But if this is so, and if it is also true that a large stone is moved with eight degrees of speed, for example, and a smaller one with four, then joining both together their composite will be moved with a speed less than eight degrees."[33] Second, w.r.t. imaginings, these moveables are the imagined balls in James R. Brown's account of the preceding: "We are then asked to imagine that a heavy cannon ball is attached to a light musket ball."[34] Third, w.r.t. suppositions, consider the train ones in Einstein's thought experiment about the relativity of simultaneity: "Up to now our considerations have been referred to a particular body of reference, which we have styled a 'railway embankment'. We suppose a very long train traveling along the rails with the constant velocity v and in the direction indicated in Fig. 1."[35] Fourth, w.r.t. imagining yourself in a situation, consider Judith Jarvis Thomson's Violinist Thought Experiment: "But now let me ask you to imagine this. You wake up in the morning and find yourself back to back in bed with an unconscious violinist…"[36] Alternately, we may check our answer against various prominent writings in the secondary literature. Again, unsurprisingly, we would find support. Consider some examples:

- Mach: "All of them [i.e. thought experiments] imagine conditions"[37]
- Duhem: "expérience fictive" (fictional experiment); "le physicien imagine une expèrience qui, si elle était exécutée" (the physicist imagines an experiment that, were it executed)[38]
- Kuhn: "the situation imagined in a thought experiment"[39]
- Sorensen: "The audience [of a thought experiment] is being invited to believe that contemplation of the [experimental] design [without executing it] justifies an answer to the question or (more rarely) justifiably raises its question."[40]
- Brown: What, in a thought experiment, one is "asked to imagine,"[41] one is to "Suppose,"[42] or one is to "imagine"[43]
- Norton: "hypothetical or counterfactual states of affairs"[44] and "imagining what the world might do if we were to manipulate it in this way or that"[45]
- Miščević: "the thought experimenter is allowed to… freely imagine things and then rearrange these imagined items."[46]
- Nersessian: "a dynamical model in the mind by the scientist who imagines a sequence of events and processes"[47]

33. Galilei, *Two New Sciences: Including Centers of Gravity and Force of Percussion*, 66–67.
34. Brown, *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, 1.
35. Einstein, *Relativity: The Special and the General Theory*, 29.
36. Thomson, "A Defense of Abortion," 48. For discussion, see 2.3.2.
37. Mach, "On Thought Experiments," 136.
38. Duhem, *La Théorie Physique: Son Objet, Sa Structure*, 163.
39. Kuhn, "A Function for Thought Experiments," 241.
40. Sorensen, *Thought Experiments*, 206.
41. Brown, *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, 1.
42. Brown, 3.
43. Brown, 8.
44. Norton, "Thought Experiments in Einstein's Work," 129.
45. Norton, "Why Thought Experiments do Not Transcend Empiricism," 44.
46. Miščević, "Mental Models and Thought Experiments," 215.
47. Nersessian, "In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling," 292.

Now, among those cases which we're unreflectively inclined to call "thought experiments," we can also find intermediate cases, ones which "do and do not" involve imaginings.[48] Consider the famous thought experiment in Thomas Nagel's "What Is It Like to Be a Bat?"[49] In it, among other things, we are supposed to be unable to imagine what it's like to perceive the world as a bat does with sonar, which inability is to bring out how objective and subjective conceptions of something differ—which difference, in turn, bears on, e.g., the mind-body problem, especially the great difficulty of giving a full explanation of mental, subjective phenomena in purely physical, objective terms. Ask: Does this thought experiment involve imaginings? At a glance, no, since, in it, arguably, one does not imagine anything, and also cannot—but, on reflection, also yes, since, in it, arguably, one fails to imagine. It's an intermediate case. By contrast, no such intermediacy arises for our paradigms above, since they have in them both imaginings—cannon balls, trains, violinists, etc.—and so too directions to imagine.

Here, we see a not common feature—and, also, to be sure, a not not-common one—that explains our unreflective concept of a thought experiment; and, it is a feature that, no less than any other, would belong to a successful definition that explains the concept by appeal to a unique conjunctive set of conditions—were such success possible.

But, one may worry, is there really such a thought experiment in Nagel's paper? For, if not, I've found no intermediate case. I'll develop and then respond to this worry in three ways. By the way, in so doing, I'll be responding to worries about error and stipulation, which I argued above present difficulties for the Gendler-Häggqvist argument; but, here, having begun with paradigms and their features, arguably, we'll be in a good position to deal with them.

First, is it really a thought experiment? I think it goes without saying that neither am I mistaken nor have I stipulated a definition under which it so counts; but, to be sure, consider what Nagel aims to do with the bat example: "To illustrate the connection between subjectivity and a point of view, and to make evident the importance of subjective features, it will help to explore the matter in relation to an example that brings out clearly the divergence between the two types of conception, subjective and objective."[50] That is, he uses an example, the bat one, to explore something abstract and, thereby, bring out a conceptual difference, which will, in turn, help him illustrate an abstract relation and make evident the importance of certain features. This, on the face of it, justifies calling Nagel's exploration—which aims, by means of example, at knowledge and understanding—a thought experiment.

Second, does the thought experiment really involve no objects of the imagination? Suppose, for reductio, that, to the contrary, it does. The only candidates seem to be certain examples of imagining that we are bats, e.g., flapping our arms or changing our bodies into those of bats, which, reading Nagel's paper, we may well imagine kinematically or pictorially. Could we be cued to imagine such things or is Nagel reporting his imaginings? Perhaps we're supposed to try to imagine, as we think Nagel did, what it's like to be a bat and fail, the failure showing us that we cannot know it, or something like this. No? No. For Nagel cues no such imagining, the examples serving merely to illustrate what a human might try to do, and there's no textual reason to think he's reporting his imaginings. The reason he gives for believing we humans cannot imagine doesn't come from one trying it out but from an in principle argument. Namely, in short, "if I try to imagine [what it is like to be a bat], I am restricted to the resources of my own mind, and those resources are inadequate to the task."[51] More specifically, he assumes a Humean conception of imagination as experience

48. Cf. *PI* §122.
49. Nagel, "What Is It Like To Be a Bat?"
50. Nagel, 437–8.
51. Nagel, 439.

rearrangement and then argues, in light of certain physiological differences between bats and humans, that we humans cannot rearrange, i.e., imagine, our experience such that we could perceive what a bat perceives via sonar as a bat perceives it. To be sure, we may often call a thought experiment, inspired by this paper, and that involves imaginings, "Nagel's Bat Thought Experiment," but their existence clearly raises no real difficulty.

Third, does Nagel's thought experiment really not involve imaginings, given that "involves imaginings" may perhaps mean "involves exercises of our capacity to have imaginings" and all thought experiments might involve that capacity? It really doesn't. After all, we do not have any such specification in mind when we appeal to imaginings to explain what thought experiments are. We appeal to imagined cannon balls, trains, violinists and so on. To be sure, to argue that it must really refer only to exercises of the capacity because, otherwise, there's no common property, is, of course, to beg the question.

In sum—since Nagel's bat thought experiment doesn't involve imaginings as many paradigms do, but involving them does, as a matter of course, explain, at least in part, why we unreflectively call something "a thought experiment"—we cannot explain why we call them so by appeal only to common properties. So we cannot give a successful sharp definition. By analogy, we cannot trace the box's ragged top sharply.

2.2.2 When to Say What Thought Experiments Are

We've now seen that our unreflective concept of a thought experiment satisfies the no-sharp-definition condition, from §1.1.2, on having a family resemblance character. To see that it satisfies the other, affinities-explained one, next section I'll survey a swath of these affinities. To give this some context in the literature, I will, in this section, contrast a position I thereby take with a couple others.

To begin, ask: When should we "define" thought experiments? Consider two answers, James R. Brown's and Sören Häggqvist's. Brown holds, in short, that we should examine examples, investigate thoroughly, and then define them. That is, first, he says, we can talk about thought experiments—even though we don't have a precise definition of them—because "[w]e know them when we see them."[52] That is, I take it, we recognize them as we know people, by their familiar faces, at a glance. Second, he adds, some of what should go into the definition is obvious, some not, and, to specify all that should go in, we should not stipulate a definition but, instead, argue and debate and, with luck, or else ideally, resolve on one at the inquiry's end. Finally, our best way, now, to understand what they are consists in looking at many examples. Häggqvist disagrees. He thinks that we should define them now. For they're greatly heterogeneous. Specifically, he argues, the great many differences between them prevent both "meaningful debate" and acceptable theories of their nature and workings, i.e., ones not "eclectic and gerrymandered"—unless we stipulate a definition; thus, without stipulating one, it's neither possible to investigate their nature and workings nor to define them well, and we should stipulate now—not aim to define them, as Brown thinks we should, at inquiry's end.[53]

I want to raise a similar objection to Brown's position, in order to differentiate it from my own. Before raising it, so as make clear what separates my position from Häggqvist's, I'll explain the difference. To this end, I'll make two objections to his position, after I fill it out a little.

To do so, ask: Is a satisfactory stipulation possible? Häggqvist, after the above disagreement, goes on, first, to stipulate that the expression "thought experiments" means "hypothetical examples used as tests of theories" and, second, to argue, in effect, that this stipulation supports investigation, insofar as it underwrites meaningful debate about thought experiments. Specifically, he argues (i) that this stipulation doesn't fail to count as such certain cases generally accepted to be thought experiments and (ii) that it is frequently an

52. Brown, "Why Thought Experiments Trascend Empiricism," 25.
53. Häggqvist, "A Model for Thought Experiments," 58.

intuitive explication.[54] To be sure, this argument isn't a quixotic attempt to justify a new usage by appeal to it being old. For we should not read him as proposing a new definition but, instead, as making a new attempt to have many of us define as some of us already do.

My two objections, in short, are, first, that Häggqvist has an unjustified premise about heterogeneity, and, second, that he makes a weak inference from that premise. Here are the objections in detail. As I argued above, he, ignoring stipulation and error, doesn't establish the *great* heterogeneity on which his argument depends. To be sure, the similar argument I give doesn't either, nor does it aim to, although it does leave open the possibility. But, even then—i.e., even if great heterogeneity were established—why stipulate instead of merely specifying which kind of thought experiments are in question? That is, second, it's not clear that the heterogeneity justifies going beyond saying, for example, "I want to talk only about thought experiments which are hypothetical examples used to test theories." Now, to defend these objections, I'll respond to two possible replies. First, if the heterogeneity were so great that, at least characteristically, we could not recognize thought experiments when we saw them, then we'd have difficulty talking about and so specifying the kind of them that we want to talk about, which may justify taking the leap to stipulation. But this reply isn't open to Häggqvist. To be sure, he rejects Brown's argument that, since we so recognize them, "we need not bother with defining" them at our inquiry's outset;[55] however, he doesn't object to the recognition premise itself. And he shouldn't. That is, to support his stipulation, he appeals to how it captures certain cases generally accepted in the literature as thought experiments, and to object that heterogeneity prevents such recognition would undermine the appeal by calling this general acceptance into question. Here is the second reply. Since, as I pointed out, we can think of Häggqvist's stipulation as aiming to reflect a use already present in the literature, it isn't anything beyond specifying kinds of thought experiment, but is equally drawing our attention to what's already there. Nobody, for example, need change how they use the term "thought experiment" on the strength of his arguments; one is only supposed to recognize a certain usage. But what's already there may not be stipulation at all. Consider his two supporting examples. First, he quotes Tamar Gendler: "I will assume that to perform a thought experiment is to reason about an imaginary scenario with the aim of confirming or disconfirming some hypothesis or theory…"[56] Is this stipulation and, specifically, explication? For all that's been said, no, since, on the face of it, she assumes a real definition, and nowhere is it indicated that she means to replace vagueness with sharp lines. Second, of Timothy Williamson, Häggqvist points out that he "talks of 'the use of imaginary counterexamples supposedly to refute philosophical analyses or theories', a practice discussed… under the rubric 'Thought Experiments.'"[57] Again, for all that's said, this isn't stipulation, since it could just as well be assuming a real definition or even specifying a kind of thought experiment.

In light of these objections, here is how the position I adopt differs from Häggqvist's. I don't think we've good reasons to stipulate a definition of thought experiments now instead of waiting till the end of inquiry. For we've little reason to think thought experiments so motley that, otherwise, we cannot meaningfully debate their nature and workings. At this point, we may want to accept Brown's position and shoot, instead, for a definition at our inquiry's end. As I said, my position isn't his, and, to explain how mine differs, let us turn to my objections, those which resemble Häggqvist's but are now easily distinguishable from them.

As Häggqvist did, let us argue against the view that we ought to work toward a definition at our inquiry's end and no earlier—in light of our capacity to recognize them when we see them, and so talk about them without a definition—and of our ability to understand them via examples. However, unlike him, since I do not

54. Häggqvist, "A Model for Thought Experiments," 58–60.
55. Häggqvist, 58.
56. Häggqvist, 60.
57. Häggqvist, 60.

hold thought experiment to be *greatly* heterogenous, I'll argue against a large part of a *much too strong* interpretation of the view. That done, I'll argue against a small part of a moderate interpretation of it.

Let us begin with the recognition claim, too strongly interpreted. It's false if taken to mean: necessarily, if what one perceives is a thought experiment, one knows that it is, and, if it isn't, that it isn't. After all, we can come to know that what we perceive is a thought experiment, or isn't as the case may be. For example, we can—intelligibly and honestly—ask, "Are Nagel's bat considerations a thought experiment?" And, to be sure, we might learn from the answer, e.g., "Yes, Nagel's considerations are one, even if they don't involve objects of the imagination." Alternately, it makes sense to ask sincerely, e.g., "What about Orwell's *1984*?" And one may correctly answer uncertainly, e.g., "Straight off, I'm not sure, since, although it's a work of the imagination and makes a political point, it's so long, and it's art." In this light, the recognition claim, so interpreted, doesn't sit well with certain strong denials that early definition is important, e.g., that, luck aside, it's impossible, before the end of inquiry, for a definition of "thought experiment" to be important. After all, we could use a definition, before our inquiry's end, to work out whether, or under what conditions, cases like Nagel's bat considerations, or Orwell's great novel, are thought experiments—and so whether they belong to our subject matter, or under what conditions they might.

I've objected, now, to the too strong interpretation of Brown's position. Here is part of a moderate one which I accept, but which Häggqvist, because of the inference in it, does not: Characteristically, we recognize thought experiments when we see them; so, in a large class of cases, we can talk about them without any need for a definition; also, often, we can better understand them via examples; and, for many purposes, before our inquiry's idealized end, defining "thought experiment" isn't important. Here is the part, of a moderate Brownian position, that, like Häggqvist, I do not accept: that we should aim and hope for a definition at the end of inquiry, idealized or otherwise. I do not because we've little reason to hoist this part aboard, given the rest.

To set up a reply to my objection, let me point out that the kind of definition at issue is the one Brown thinks it's not now important to give—namely, a "sharp" one—that is, at the very least, one which draws a line that completely divides what is from what is not a thought experiment, as Häggqvist's stipulation aims to do. It's of this kind because we cannot aim and hope for the alternative, since, as we saw last section, we already give them. Now, one may reply, we should aim and hope for a precise definition, since a complete theory of thought experiments should include one—one like "water = H_2O." Thought experiments, however, are, at least, fairly disparate, as we saw last section, and so, first, the idealized sharp definition would, by our current lights, be a stipulation. Second, for the good of such a theory, we have little reason to stipulate instead of merely specifying which phenomena we want to talk about.

To be clear, I mean my claims about how disparate thought experiments are and what we recognize to be one to be true at present. I allow, for example, that we might later on recognize *1984* to be a clear case of a thought experiment. This follows from the family resemblance account of our concept of a thought experiment, for which I'm arguing this section. To be sure, as we saw above in §1.1.2, we can extend such concepts, as it were, spinning fibre upon fibre.

To sum up my position: When, if not now as Häggqvist recommends, and if not at the end of inquiry, as Brown does, should we shoot for a "precise definition"? Never. This is not to deny that, either now or later, we should give an "imprecise," or especially a non-stipulative, "definition." Rather, I think we should give one for certain purposes, e.g., as we'll do next chapter, to extend our investigation into what we do not now recognize either as, or as not, a thought experiment or else to characterize the nature of our concept of a thought experiment. This characterization we will now begin to give.

2.2.3 Satisfying the Affinities-Explain Condition

The "definition" I'll give is an account of certain affinities. They are ones that, as you'll see, partly explain our unreflective concept of a thought experiment. In light of this account, we will see that this concept satisfies the other condition—i.e., the affinities-explain-it one on having a family resemblance character.

Three Worries

Before I give the account, I'll try to assuage three worries about affinities explaining the concept. The worries concern the universality of similarity, recognition learning, and a criticism of Kripke's.

First, since everything is like and unlike everything else in some respect, isn't any such account committed to the view that we recognize everything as a thought experiment? And as not one? And as an intermediate case?[58] No, because, as we'll see, the account appeals not to similarities and differences simpliciter but only to particular ones.

Second, if commonalities alone don't explain the concept, how could we possibly learn to recognize thought experiments? By analogy, a child does not—if not by doing such things as seeing, repeatedly and in various places or picture books, *a common* coloured shape called "horse," instead of "goat" or "cow"—learn to recognize horses.[59] It's learned via affinities. That is, often, we do learn to recognize thought experiments in virtue of seeing "local commonalities," e.g., those between certain pairs of thought experiments; and, we often appeal to such "commonalities" when explaining what thought experiments are; but, these commonalities need not be global, i.e., can be mere similarities, affinities. To be sure, this is neither to deny that global commonalities ever explain nor that, if they ever do, their doing so poses any trouble for my account. After all, in light of §1.1.2, it's compatible with the account that commonalities—but not only commonalities—do so explain.[60]

To render this response more plausible, let us develop its account of learning to recognize thought experiments. Consider three kinds of learning: via inadvertent-, intentional- and non-teaching. First, if inadvertently taught, the similarities by which we learn may be those salient ones among well-known, named historical examples that we've been exposed to in, say, our physics or philosophy classes, such as those discussed above in §2.1.3. Second, if recognition isn't inadvertently but intentionally taught, the similarities may be the ones appealed to across often read philosophical explanations of what thought experiments are, also as discussed above in §2.1.3. Third, if it's not taught, they may be those which we, as a matter of course, work out by ourselves using a model which the term "thought experiment" suggests, as discussed in §1.2.1.

But how could we learn "non-recognition" and "neither-recognition"? Possibly in the same way. That is, first, to recognize what is not a thought experiment, one may simply learn to recognize what is so, as in the three ways just mentioned. After all, to explain why something isn't one, we can point out certain missing similarities—i.e., certain absent ones by which we recognize thought experiments. For example, if asked whether your water bottle, there on the table, is one, to explain why not, you might point out that it lacks relevant properties, such as being an imaginable or being for gaining knowledge or understanding. Second, to learn to recognize that something neither is nor isn't a thought experiment, we may again simply learn to recognize them as above. After all, to explain that something is a borderline case of them, we may point out both certain similarities as well as a certain lack of them. For example, in light of the next chapter, we may

58. Cf. the criticism of (i) anti-definition family resemblance approaches to art and (ii) similar resemblance-to-paradigm accounts, respectively, in Carroll, "Introduction," 10–11 and Gaut, "'Art' As a Cluster Concept," 25–26.

59. Cf. *PI* §72.

60. Cf. Berys Gaut's "cluster concept" account of art on which being an artifact is a necessary condition on anything whatever counting as art (Gaut, 29).

recognize certain stories in works of literary fiction, such as Orwell's *1984*, to be borderline cases of them, since they involve uses of the imagination to make a political point, but they differ in how imagination functions, how free we are to interpret, and how complex they may be.

Here is the third and final worry. To explain it, take Searle's cluster concept account of proper names, on which, very roughly, a proper name such as "Aristotle" names Aristotle in virtue of sufficiently many unspecified descriptive statements being true of him—not necessarily all of them.[61] Also, take Kripke's criticisms of such accounts and of cluster concept accounts of natural kind terms like "gold." For him, such terms are "rigid designators," like the following: "Dartmouth," which names without a thought to the Dart's mouth; "Holy Roman Empire," which doesn't name in virtue of that institution being either holy, Roman, or an empire, since it (suppose) isn't any of them; and, "Moses," which names the man even if every one of the identifying explanations we might give, e.g., every Biblical description of him that we might appeal to, is false.[62] That is, proper names and natural kind terms, contra cluster concept accounts of them, don't refer in virtue of satisfying description-like criteria. Now, here is the worry. Since affinities are description-like, doesn't Kripke's criticism apply to the account I want to give?

Berys Gaut, responding to such a worry about his cluster concept account of art, argues that the criticism does not apply. To this end, he appeals, first, to the term "art" being neither a proper name nor a natural kind term and, second, to the inapplicability of the concept if none of its criteria are satisfied.[63] My response partly apes Gaut's. That is, I'll argue that the criticism applies only to a limited extent, since, several qualifications aside, expressions like "thought experiment" often aren't either proper names or natural kind terms.

We begin with proper names. More specifically, we begin with familiar ones, i.e., with the use of "thought experiment" in proper names, such as "Galileo's Falling Bodies Thought Experiment." We often use these names as if they were definite descriptions. For example, this last name we often use as we would the expression, "the one by Galileo about falling bodies," to figure out which one someone means. We use certain similar proper names, ones without "thought experiment," similarly. For example, we often use "Thomson's Violinist" and "Mary the Color Scientist" similarly—i.e., use them as we would "the one by Thomson about the violinist" and "the one about Mary the Colour Scientist." That is, we use such expressions to do such things as call certain thought experiments to mind, find passages expressing them, make claims about this or that one, and so on. Now, insofar as a family resemblance account can explain these particular uses, the Kripkean criticism doesn't apply to it. These uses, however, aren't the only ones. We also often use the name such that it refers even if it is, as it were, false, like "Holy Roman Empire." We do so, for instance, when we use the name to refer to whatever others do, e.g., to what scholar so-and-so does. So used, an expression like "Galileo's Falling Bodies Thought Experiment" refers even if it turns out that it's not by Galileo but Schmalileo, from whom the former stole it, as in Kripke's well-known "Gödel" case. To a family resemblance account of these other uses, the Kripkean criticism does apply. Even here, however, it applies only to the whole name, not to "thought experiment" in particular. To be sure, "Galileo" is part of such a name and the criticism would apply to a family resemblance account of it; however, it's usually a proper name in its own right. In sum, the criticism applies only to a limited extent to family resemblance accounts of such names so used.

These uses, moreover, are, so far as I can tell, the only matter of course ones. By contrast, someone might use "Thought Experiment" as both a proper name and honorific. It might refer to a particular thought experiment we believe is the greatest, much like medieval scholars used "The Philosopher" for Aristotle. This use, however, if extant, would hardly count as a normal one. So, by my lights, then, so far as proper names go,

61. Searle, "Proper Names."
62. Kripke, *Naming and Necessity*.
63. Gaut, "'Art' As a Cluster Concept," 26–30.

the Kripkean criticism applies only limitedly to a family resemblance account of matter of course uses of terms like "thought experiment."

It also applies only limitedly so far as natural kind terms go. I'll give two arguments for this claim.

One is that expressions like "thought experiment," normally used, are often not such terms, since we often cannot imagine their bearers without the relevant explanatory properties. To explain, we can imagine gold being green instead of yellow, if greenly lit, or light instead of heavy, if in space, or a gas instead of a solid, if sublimated, and so on—but not as an element with anything other than 79 protons in its nucleus. If we think we imagine it having any other atomic number, we're simply imagining another element. More generally, a word is a natural kind term—insofar as Kripke's criticism goes—*only if* we can imagine its bearer without accidental properties but cannot do so without its essential ones. Now, terms like "thought experiment" often don't satisfy this condition. After all, (i) we can imagine thought experiments without some of the properties by which we normally explain what they are, and (ii) we cannot imagine them without all of these properties. That is, either (a) the condition doesn't always apply, since these explanatory properties are neither accidental nor essential, or (b) it isn't always satisfied, since we can't imagine them without the "accidental" properties and can imagine them without some "essential" ones. So, by my lights, the Kripkean criticism applies only limitedly to family resemblance accounts of normal uses of terms like "thought experiment" as natural kind terms.

Here is the other argument that this is so. Suppose that water is under investigation, that we do not know its chemical structure, but that we do have the Periodic Table. In this case, the term "water" may rightly be called a natural kind term, even if we cannot now imagine water absent the accidental properties so far observed in our samples. We sometimes investigate thought experiments similarly. For example, we posit common properties, such as being a mental model, to explain a class of phenomena we call "thought experiments," and we evaluate our hypothesis in light of empirical evidence. We may even aim at an ideal, i.e., to establish a sharp definition. In such cases, as we could "water," we can call an expression like "thought experiment" a "natural kind term," even if we cannot now imagine its bearers without all the properties by which we now explain what they are. To a family resemblance account of such "natural kind terms," the Kripkean criticism doesn't apply, since the terms don't satisfy the above condition on being of that kind. This, clearly, also limits its applicability.

To sum up my response to this third and final worry, the Kripkean criticism doesn't fully apply to a family resemblance account of expressions like "thought experiment." It doesn't because, certain qualifications aside, these expressions, normally used, are often neither proper names nor natural kind terms.

Giving the Account

Worries assuaged, let us turn to the account, which consists in two sketches. The first concerns affinities that we weave using a comparison to an ideal which the term "thought experiment" suggests, as we saw in §1.2.1. The outcome is a familiar, rich and compelling conception of thought experiments—as experiments in thought. This sketch sheds light on overlapping similarities at different levels of generality. To see how they also criss-cross, i.e., their disorder—in line with §1.1.2—we turn to the second sketch. There we see other, independent affinities, which we weave, largely, from named historical examples.

To begin the survey's first sketch, in the literature, as we've seen, we often see thought experiments characterized by comparison to experiments. Again, terms like "thought experiment" invite this form of description.

This may strike us as extraordinary: Word reveals world! What providence! Well, at least it can strike us so if we forget how we learnt the terms. Often, they invite us, while trying to understand them, to compare them

to words like "experiment." For example, we contrast them with "physical experiment," or we see "thought" as adjective, or adverb, modifying "experiment" as noun, or as verb. That is, a learning comparison becomes, no surprise, a describing comparison.

In what follows, I'll sketch some of this learning. That is, I'll sketch certain affinities that, using experiments as a comparison, we select and weave together. We are to see here affinities, or family resemblances, as opposed to commonalities alone, that partially explain our unreflective concept of a thought experiment. To this end, I'll draw on and recast descriptions of affinities given in an earlier work,[64] doing so under three heads: "Imagination," "Action," and "Aim." I recast them variously, but, perhaps most importantly, I do not, in light of eschewing theoretical posits in §1.2.2.1, say here that if any description doesn't seem plausible in its own right, consider it a conjecture.

Also, for illustrative purposes in what follows, here is the gist of Einstein's Elevator Thought Experiment. Imagine an elevator accelerating upwards—deep in space. A scientist inside observes that balls, when released, move downward—and do so just as if they were falling due to gravity in an elevator hanging near the Earth's surface. It would be a mistake to conclude that the elevator is at rest, no? No! For that's an equally justifiable way to regard the situation. This equality, moreover, grounds our taking natural laws to be the same across constantly accelerating reference frames, not only inertial ones—i.e., grounds the Equivalence Principle in Einstein's General Theory of Relativity.[65]

Imagination If we model thought experiments on experiments, and notice our imaginings of such things as elevators in space as well as words like "imagine," "if" and "suppose," we're inclined to treat these imaginings, or hypotheticals, as in the one much like observations are in the other—that is, treat them as if they function like observations. This language can mislead us, e.g., lead us to think imaginings are essential to thought experiments as observations experiments. Here, however, I'm concerned with how we use the model to weave our concept of a thought experiment—i.e., how imaginings come to determine and explain the concept. In particular, while trying to work out what thought experiments are, we tend to look into the term "thought experiment," see in it a structure resembling "experiment in thought," and then—noticing our imaginings, e.g., recalling our mental elevator image—describe them as in thought experiments, doing so much like we describe observations as in experiments. Our descriptions, which often borrow words from psychology and logic, take such forms as: thought experiments involve imaginings, or include an imaginable, or have in them an imaginary case or scenario, or are visualizable, and so on; alternately, they involve asking "what if," or counterfactuals, or hypotheticals, or entertaining a proposition either without regard to its truth-value or else alongside the belief that it is false, and so on.

Also, as they should *qua* affinities, these similarities overlap and criss-cross at different levels of generality. For example, at a general level, Einstein's Elevator involves the imagination like both Thomson's Violinist, as we'll see in §2.3.2, and Galileo's Falling Bodies one, but unlike Nagel's Bat; and, at a specific level, it involves something visualizable, also like these two, but unlike P.F. Strawson's purely auditory thought experiment,[66] Thomas Reid's purely tactile one,[67] and Berkeley's in which we're to fail to imagine certain unperceived objects.[68]

64. McComb, "Thought Experiment, Definition, and Literary Fiction," 209.
65. Einstein, *Relativity: The Special and the General Theory*, 75–79 and Einstein and Infeld, "The Evolution of Physics," 230–235.
66. Strawson, *Individuals: An Essay in Descriptive Metaphysics*.
67. Reid, *An Inquiry into the Human Mind on the Principles of Common Sense*, 65–67.
68. Berkeley, *Principles of Human Knowledge and Three Dialogues Between Hylas and Philonous*, 60–61.

Action If we weave our concept of a thought experiment using experiments as an ideal, we're inclined both to notice certain goings on, like a scientist working in an accelerating elevator, and to treat these events as actions in the one much like manipulations in the other. For instance, we often speak of Einstein carrying out his elevator thought experiment, or of getting his reader to do so—and, on the off chance we're asked to elaborate on this quasi-experimental procedure, of not only having an imagining but imaginatively accelerating the elevator, having the scientist inside test hypotheses, observing the consequences, and so on.

Again, three points: First, notice that these similarities aren't commonalities, since some thought experimental situations do not permit such manipulation, e.g., the static spheres in Black's thought experiment discussed below, in §2.3.1. To be sure, we can call successively combining properties to make up a situation "mental manipulation," or "building a mental model," and call the conclusions we draw from reasonings about the situation "results." Nevertheless, the differences in what we do are obvious. Put another way, both sorts of action are similar in general but different in particulars. That is, second, the similarities vary at different levels of generality. Third, in the above descriptions, these varying similarities evidently overlap with the preceding imagination ones, insofar as they're actions upon them.

Aim Modelling thought experiments on experiments—and noticing, for example, Einstein's final assertion above, the one about the elevator scientist's equivalent justification and its yielding good grounds for the Equivalence Principle—we're inclined to treat it as in thought experiments much like results are in an experiment. Treating such things so, we often describe them as results or outcomes—or, if comparing to arguments, as conclusions—and say they consist in a piece of knowledge, justification for a belief, understanding, and so on. We may also specify that this outcome or conclusion concerns more than the result's source. And, so, we also separate results from other parts of a thought experiment. These parts include the source of the results, e.g., a case or scenario. They may also include the relation between source and purported result, e.g., an intuition, or an inference, or induction, or confirmation, and so on.

Again, three points: First, these similarities do not amount to commonalities, since not all thought experiments have a distinct result. The Clock in a Box thought experiment doesn't, for example, since it may have one or another result, on which see §3.4.1 below. Indeed, we often use such names to pick out a certain imaginary situation apart from what it purportedly shows. Another example: we use "Nozick's Experience Machine" to refer to that situation in which you are to choose whether or not to plug into the up-to-you experience-giving machine but neither to Nozick's argument for his choice nor to that choice as the result.[69] Second, the similarities occur at different levels of generality, since two thought experiments may be similar insofar as they have a result but different insofar as the result is of a particular kind, e.g., a justified belief or a clarified idea. Third, these similarities overlap. We see this when we've woven this strand into either of the preceding two, thereby arriving at our rich and compelling conception of thought experiments as those performed in mentis. We can outline, for example, three stages in a thought experiment, e.g., imagining a situation, varying variables in that situation, and, from the outcomes of these varyings, achieving results.

This first sketch may give the impression that the affinities overlap in an orderly way, but they also criss-cross. To bring this out, and to transition into next chapter's ideas, turn to the second sketch. It concerns two similarities woven in using comparisons of another kind—i.e., to well-known named historical examples. The two similarities are: first, being surveyable, as fables tend to be; and second, having a particular point, somewhat like a sonnet's final couplet.

To begin, recall the above forms of proper names for thought experiments. They resemble *So-and-so's*

69. Nozick, *Anarchy, State, and Utopia*, 42–45.

such-and-such thought experiment. There is, for instance, "Einstein's Elevator thought experiment" but also "EPR," which has the form "So-and-So," and "Mary the Colour Scientist," which has the form "Such-and-Such." What remains fairly stable across our uses of such names are author attribution and memorable-feature reminding. This makes salient two sorts of similarities, which overlap across the names' bearers—namely, those concerning the author's hand and easily remembered features. These stable and salient similarities, we're inclined, unsurprisingly, to appeal to them when, looking to well-known historical examples, we explain what thought experiments are. Under these two similarity kinds fall the two I said I'd consider.

Fixity of Point When we explain the nature of a particular historical thought experiment, we normally appeal to its point, one proper to it and external to its case or situation. For example, Einstein specifies the point of his elevator thought experiment, which lies, as in induction, beyond the included situation or case, i.e., concerns the Equivalence Principle, not just an elevator or scientist; and, you do not understand it, as seen in your explanation of it, if you do not know that it "has that point," in this sense, even if you disagree with it or "carry out the thought experiment" and make another point. That is, one widely shared explanatory property of being a thought experiment is having a fixed point, not an open-ended one. We will discuss this in some detail next chapter, in §3.4.1.

For now, two brief remarks. First, I don't deny that "the point changes" from the context of understanding an individual thought experiment to those of evaluating or retooling it. For example, one may argue that its "real point" isn't its author's but a certain other. Alternately, one may explain how "it" can be reinterpreted to make another point. Second, to bring out the criss-crossing, or messiness, contrast this explanatory property with those above titled "Aim." This one is a particular point. Using experiments as a comparison, we may also weave in a particular point; however, we may also weave in any old point or outcome—e.g., whatever one who carried it out arrives at—or no point beyond certain phenomena—e.g., as in experiments that aim merely to generate them.

Suveyability As for the second similarity, when we explain what a thought experiment is we often appeal to a feature of its imaginary case or situation, which is connected to its surveyability—i.e., to it being (i) short and simple enough to take in at a glance, (ii) memorable, and (iii) reproducible.[70] Our use of thought experiment names, sketched above, moreover, depends upon it usually being surveyable. More to the point, we do not often recognize something as a thought experiment unless it's short and simple, much like an Aesopean fable. To extend the analogy, were a tale, told about talking creatures ending in a moral, to take several hours, the story so complicated that it would be impossible either to take it in at a glance, to remember, or to reproduce it, we wouldn't recognize it as a fable, at least without, e.g., some thought and effort. We'll discuss this too in some detail next chapter, in §3.4.2.

Again, two remarks. First, granted, we're inclined to say that thought experiments may be any length. And we may think this shows shortness, and surveyability more generally, not to be an explanatory property. But whence the inclination? If from modelling on experiments, as above, all that might show is conflict between models. The objection, undeveloped at least, doesn't stick. Second, this explanatory property may criss-cross instead of overlapping regularly, since it arises independently from those in the first sketch. After all, assuming experiments generally aren't surveyable, the above modelling on them provides for no such explanatory property.

This completes the second sketch and, with it, the broader survey, in which we see affinities that explain,

70. Cf. Wittgenstein, *Remarks on the Foundations of Mathematics*, II.

albeit partially, our unreflective concept of a thought experiment. This done, we also see that the concept satisfies more than the no-sharp-definition condition on having a family resemblance character. It also satisfies the other, affinities-explain one. Or so I've argued in this section.

2.3 On Thought Experiments without Imaginings

A central claim so far has been, in short, that imaginings aren't essential. That is, having them explains our unreflective concept of a thought experiment—but it is not a commonality. The claim hardly rings true. Why not? In part, it may well be because we overlook a rule for learning the concept—an imaginings-free, stop-sign-like one for the expression "It's a thought experiment." To support this contention, in line with §1.1.1, I draw attention to that rule by comparing language games. I do so preliminarily, in §2.3.1, and then in further detail, in §2.3.2.

2.3.1 Preliminary Comparisons

A Sequence of Dialogues

Identifying the Point

PROFESSOR: Listen.

> Isn't it logically possible that the universe should have contained nothing but two exactly similar spheres? We might suppose that each was made of chemically pure iron, had a diameter of one mile, that they had the same temperature, colour, and so on, and that nothing else existed. Then every quality and relational characteristic of the one would also be a property of the other. Now if what I am describing is logically possible, it is not impossible for two things to have all their properties in common. This seems to me [that is, to Max Black] to refute the Principle.[71]

What's the point?

STUDENT: What it says in that last sentence, you know, about the Principle being refuted.

Understanding Black's Point Professor then tests student's comprehension:

P: Can you rephrase Black's point?

S: Sure. It's that what he described refutes a principle, the Principle of the Identity of Indiscernibles. Oh and he hedges the claim. He says that it *seems to him* to refute it.

P: Nicely put. And what does he think he described?

S: Those, you know, two spheres that have all the same properties but are still two different things.

Test passed. Alternately, he fails:

S: Rephrase it? OK umm what's the Principle?

P: It's the Principle of the Identity of Indiscernibles, Leibniz's, from yesterday.

S: Oh, right, so the point's that it's refuted?

P: Good but how? Refuted by what?

S: I dunno… something about temperature and colour and stuff?

71. Black, "The Identity of Indiscernibles," 156.

Judging Black's Point Professor has student evaluate the point:

P: Does it seem to *you* that what Black described refutes the Principle?
S: Yeah, I guess so.
P: So, what? He established that two different things can have *all* the same properties?… Yes? [Student nods.] All right. But don't you think that there must be some property that makes the two spheres *two*? That otherwise they'd be *one*?
S: Well, maybe they have the same properties but different matter, the two spheres, you know, like pin cushions… how they're two even if they have the same pins.

Student defends the point. Alternately, he doesn't:

P: Do you accept Black's point?
S: Can't, no. They're in different places, the two spheres.
P: You think they have different spatial properties? So what?
S: Because then they don't have all the same properties.
P: Good. So you think Black didn't describe two spheres with *identical* properties? That instead he described two spheres with different properties, different spatial ones, and that such a description doesn't refute the principle?
S: Sounds right. Yeah.

A Misunderstanding Finally, and most importantly, the student misunderstands:

P: [Reads Black's Thought Experiment]
S: But it's not true! The universe has more than two spheres in it! And no two spheres anywhere are exactly alike!
P: It's a *thought experiment*!

Or:

P: [Reads Black's Thought Experiment]
S: But it's not true that nothing but those two spheres exists!
P: You don't understand. I read you a *thought experiment*.
S: Um, so… I don't get it.
P: Look. I did read these words, "each [sphere] was made of chemically pure iron, had a diameter of one mile" and "nothing else existed," and so, you're right, more or less; I did say the universe contains nothing but the two spheres, but the words do *not* mean that the universe *really* contains nothing but the two spheres. You didn't take the passage as you should have. I read you a *thought experiment*.

The student misunderstands by "disagreeing." He may also do so by "agreeing":

P: [Reads Thought Experiment]
S: Right, that's true. The universe contains nothing but two identical spheres.
P: It's a *thought experiment*! [72]

72. Or, both misunderstand:

S: Right, that's true. The universe contains nothing but two identical spheres.

***

Black's thought experiment serves as a pedagogical tool, a philosophical watering can; with it, professor teaches student about a thought experiment. This use, by the way, need not be the only one. For example, with it, the professor also teaches Leibniz's Law.

All this teaching counts, in two ways, as a language game. First, it's a whole made up of words woven into actions. Second, we can think of the whole process—that of the student learning how to identify, understand, and judge the point or the part that is like a fable's story—as a game by which children learn their native language, e.g., one by which they learn how English words in thought experiments function.

Let us compare these language games. The results will illuminate how we use the term "thought experiment."

Comparisons

From the sequence, take three dialogues, those in which a student fails to understand.

In one, the student may understand *what* the professor read—namely, a thought experiment—but fails to understand *its point*. After all, unlike the preceding dialogue's student, he *fails* the professor's test. Now, were she, the professor, to say, "It's a *thought experiment*!"—and were the student to rightly respond, "I get that! Just not *its point*!"—she would *not* thereby get him to understand.

In the other two dialogues, in which the student misunderstands *what Black says he described*—i.e., the story, as it were—the professor *would* get him to understand—by saying "It's a *thought experiment*!"—if, roughly-speaking, she thereby stops such inept replying.

Another pass. In these two dialogues, the student affirms in one and denies in the other a proposition about Black's spheres—that they alone exist and that there are two identical ones—as if the professor had read from a cosmology textbook or *Scientific American* article entitled, "What the Universe Contains." In so doing, the student obviously misunderstands what was read. Now, were she, the concerned professor, to add, "It's a *thought experiment*!"—aiming to correct his misunderstanding—and were this addition to stop him, when responding to similar passages, from making any such affirmation or denial or the like—then she'd succeed. That is, she'd dislodge an obstacle to his understanding thought experiments like Black's.

Obstacle removed, this student—no longer inclined to disagree with such sphere-propositions, or to agree with them, or the like—might understand what was read, just as well as those students who agree or disagree with the thought experiment's point.

In sum, under certain conditions, "It's a *thought experiment*!" clears away a misunderstanding about a thought experiment—specifically, about its "story." Under others, by contrast, it does not do so—in particular, clear one away about a thought experiment's point.

Such uses of "thought experiment" are my main concern here. Before further illuminating them, consider a source of doubt.

P: No, you're wrong. It contains much else and no such spheres!

Or, the professor misunderstands by "agreeing with the student":

S: But it's not true! The universe has more than two spheres in it! And no two spheres anywhere are exactly alike!

P: I know, right?! There's *no way* it's true! What garbage we fill students' heads with.

A Source of Doubt

Again, the student who replied—"But it's not true! The universe has more than two spheres in it! And no two spheres anywhere are exactly alike!"—*misunderstood* what was read. This I've been assuming—e.g., when I claim that the misunderstanding disappears, under certain circumstances, when the professor says, "It's a *thought experiment*!" And I think it's *obvious*. But you might disagree. Your doubt might arise from the fact that—with suitable additions to the dialogue—the student doesn't misunderstand. This doubt I'll try to inflame then assuage.

To begin, certain additions to the dialogue flip what we can say about it. An example, with italicized additions:

PROFESSOR: Listen.

> Isn't it logically possible that the universe should have contained nothing but two exactly similar spheres?… This seems to me to refute the Principle.

STUDENT: But it's not true! The universe has more than two spheres in it! And no two spheres anywhere are exactly alike!

P: It's a thought experiment!

S: *Oh, right, of course! I don't know what I was thinking!*
(Aside) There really aren't such spheres, but I didn't want to look stupid. So I didn't disagree but, instead, faked agreement.

Additions made, is the misunderstanding dislodged? I expect you'd say, "No, because secretly the student didn't change."

Now, here is the doubt. No misunderstanding was dislodged, once the additions were made, because the student was secretly recalcitrant. So we cannot say the professor dislodged it. We also cannot say it *before* any line was added, that is, of the student in the original dialogue. After all, that student might have been secretly recalcitrant. This is a possible fact. Put another way, that student either secretly kept thinking the same thing or did not. The matter isn't specified.

To assuage this doubt, first, take the skeptic's premise—namely, that, in the original dialogue, the student might be secretly recalcitrant. This premise treats the original as if it were an article reporting what a talking fox once said. To see this, let us model the original dialogue on part of a newspaper report, then a fable.

The report: "RCMP said 20-year-old [X] and a 16-yearold [sic] youth are charged with second-degree murder in the death of 42-year-old [Y]."[73] Notice that what the accused did at the time of the crime either counts as second-degree murder or does not, assuming sharp relevant legal definitions. Now, had the report's next statement been, "They either committed the crime or did not," would you, upon reading it, have learnt anything from it? Your gut reaction, I expect, is something like: "No, because it goes without saying. *Of course* they did or didn't do it."

To be sure, upon reflection, you may respond differently. Wouldn't you have learnt that there are relevant sharp legal definitions? After all, you might reason, why else would the sentence be there? But you need not so respond. There may be no reason; it might be a mistake like the report's missing hyphen. Or, you might think, you'd have learnt that the newspaper's editor slipped up. But it may be that nothing of the sort occurs to you, and so again you need not have learnt anything.

Such reflection aside, you might say that, in the addition-free report, the accused might have committed the crime or, alternately, their guilt or innocence isn't specified. This proposition resembles our skeptic's initial

73. Modjeski, *Murder Victim 'Kind' and 'Caring': Family*.

premise, about the original-dialogue student being possibly recalcitrant. Now, notice that it's as if, in asserting the premise, our skeptic models the dialogue on such a report—as if the dialogue were, say, journalist discovered, written up, submitted, and published in a local newspaper. To see how this misleads, consider another model, a fable.

We might add, to the following fable, the following sentence, in italics:

> A fox happened to find a mask used for performing tragedies and, after turning it this way and that several times, she remarked, "So full of beauty, so lacking in brains!"
>
> *The fox either meant what he said or did not.*
>
> This is a saying for people to whom Fortune has granted honour and glory, while depriving them of common sense.[74]

Had this added line been in the original fable, would you, upon reading it, have learnt anything from it? I expect your gut reaction to be something like this: "I wouldn't have thought that the fox might not mean what he says—and so, yes, I'd have learnt something, in particular, that the fox might not be straightforward and honest but sly and deceitful."

To be sure, again, on reflection, you may respond differently: "Isn't every statement either meant or not meant, and don't we all know it?" But you need not respond this way—i.e., take the addition to be for saying that here the law of excluded middle holds. After all, you might instead take it, e.g., as a translation error or to be poorly written.

Now, this gut reaction differs from that to the report addition, in this way. At first glance, I take it, you thought the added "They either committed the crime or did not" did *not* tell you anything but that the other one—i.e., "The fox either meant what he said or did not"—did. That is, the other did tell you something, namely, that the fox might not mean what he said. In short, the fable addition, unlike the report one, "introduces a possible fact."

Now, we can model the dialogue on the fable and the report. In particular, we can take the dialogue's addition, which in effect says, "the student is secretly recalcitrant," to be *like* the fable's, "the fox either meant what he said or did not," and *unlike* the report's, "they did or didn't commit the crime." This modelling sheds light. We can now easily see that our skeptic—when asserting, "in the original dialogue, the student might have been secretly recalcitrant"—speaks as if the dialogue were not fable- but report-like—that is, as if the possible recalcitrance weren't something that needed to be introduced to be there but were instead something already there right ready to be described.

2.3.2 Further Comparisons

The above comparisons shed light on a use of "thought experiment"—one for stopping inept responses and so removing misunderstandings. Here, further comparisons shed further light. In particular, they shed light on an imperative form of the use. It is: *Don't do anything like this in response to anything like that.* Illuminating this use, as I said above, helps us see why denying the commonality of imaginings in thought experiments seems implausible.

These further comparisons resemble those above. That is, they will be to similar activities in similar surroundings. The surroundings are, first, a fable and, second, another thought experiment. The activities in each are other means of teaching—specifically, by reminding, correcting, and training. By analogy, you may use the phrase "It's a thought experiment!" to *remind* a student about what you read—like using "keep your

74. Gibbs, *Aesop's Fables*, Fable 550.

head down" to *remind* Golfer to do so and so lessen that awful slicing; you may use the phrase to *correct* a student's false belief that you read, say, a news report—like using "it's a putter, not a driver" to *correct* Golfer's belief that putters are for driving and so stop that ridiculous putter-driving; and, you may use the phrase to *train* a student not to make inept utterances—as timely yells curtail pet presents on the floor.

Fable Dialogue & Comparison

Consider a fable:

PARENT: [To Child] Should we read a new one?… Yes?… OK.

> **The Wolf, the Dog, and the Collar**
>
> A comfortably plump dog happened to run into a wolf. The wolf asked the dog where he had been finding enough food to get so big and fat. "It is a man," said the dog, "who gives me all this food to eat." The wolf then asked him, "And what about that bare spot on your neck?" The dog replied, "My skin has been rubbed bare by the iron collar which my master forged and placed upon my neck." The wolf then jeered at the dog and said, "Keep your luxury to yourself then! I don't want anything to do with it, if my neck will have to chafe against a chain of iron!"[75]

What the wolf just said is the moral of the story. Sometimes it's *inside* the story. Anyway, what's the wolf saying?… You don't know?… That's OK. Let's figure it out together. Would the wolf trade places with the dog?… Right. Would not. But why not? The wolf would get lots of food.… Sure, good, because of the mean collar. So what do you think the moral is?… OK, but it's not just about food. Remember it's luxury to the wolf… That's a great answer! But it's also not just about a collar. Why does the dog have one on?… Right. I wonder, is the dog *free* to run away?… All right, so what's the moral?… Good try, but, well, how about using the word "trade"?… Great! "Don't trade freedom for luxury." You got it. A good moral. Now, if *you* have freedom, should *you* trade it for luxury?… Ah! Wonderful answer. "Never!" That's exactly right.

This parent, notice, trying to inculcate the moral, continually draws the child's attention to it, the moral, as opposed, e.g., to how surreal a talking wolf is, not to mention one that jeers.

Via this attention directing, a fortunate child may avoid misunderstanding the fable's story, e.g., by not even thinking to deny that a wolf can jeer or that other surreal events really occur. Three analogies. First, teaching chess, you might repeatedly move the pawns like this, the king like that, and so on for the other pieces—and, thereby, your pupil may only move these pieces in those ways and so avoid certain illegal moves—as if sticking to a familiar path for fear of getting lost. In case this analogy misleads, second, the parent may teach the child not to misunderstand the story *inadvertently*, e.g., without saying anything like, "I will direct my munchkin's attention and so teach the little one not to misunderstand the fable in this way." Third, this teaching resembles that of an ill-written though page-turner novel, one not meant to tune up your feel for prose quality—e.g., not meant to sensitize you to pointlessly trite or wordy phrases—but which nevertheless does.

This teaching—like that using "It's a thought experiment!" above—works in none of the following three ways: by *reminding* the child, for whom the fable is new, not to misunderstand the fable's story; by correcting false beliefs about that story; or, by puppy-puddle-prevention-like training. The first comparison now made, let us turn to the second.

75. Gibbs, *Aesop's Fables*, 5.

Thought Experiment Dialogue & Comparison

In the following way, and in similar ones, we often teach or learn about a thought experiment.

TEACHER: Class, today you'll learn about something new, something called "a thought experiment." We'll have a look at a famous one. It's about a violinist. First though [click] have a look at this slide:

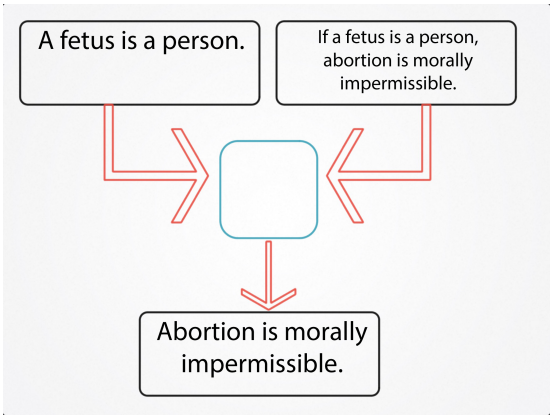

Figure 2.1: The Simple Argument

Premises go into the box. Out comes the conclusion. Basically the argument is that abortion is wrong because a fetus is a person. Familiar?… Good. [Click.] Here's another version of the argument:

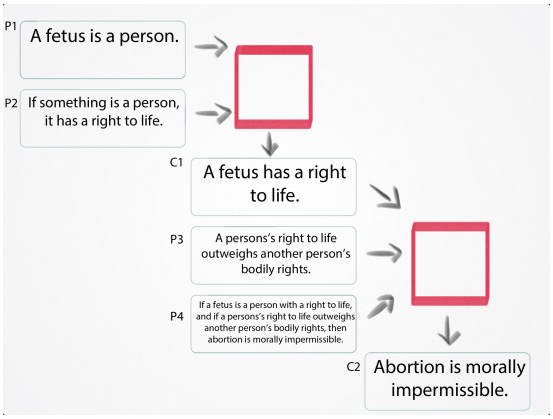

Figure 2.2: The Less Simple Argument

This version has a few extra steps between a fetus being a person and abortion being wrong, which are basically that persons have a right to life and that this right outweighs other bodily ones—and, this argument, more or less, is what the famous thought experiment is about. Here's [click] how the philosopher who made it up, Judith Jarvis Thomson, describes it, oh, and if you like, follow along as I read:

> Every person has a right to life. So the fetus has a right to life. No doubt the mother has a right to decide what shall happen in and to her body; everyone would grant that. But surely a person's right to life is stronger and more stringent than the mother's right to decide what happens in and

> to her body, and so outweighs it. So the fetus may not be killed; an abortion may not be performed.[76]

Now, this argument, Thomson thinks some important anti-abortion arguments are like it, and even that it sounds plausible, I mean, that it seems like a good argument, but she wants to cast doubt on it, assuming—and this is crucial—*assuming for the sake of argument* that a fetus is a person. Now, to cast doubt on this argument, Thomson uses her violinist thought experiment. [Click.] Here's the first bit:

> [The argument] sounds plausible. But now let me ask you to imagine this. You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own.[77]

So far so good?… Now [click] pay close attention:

> The director of the hospital now tells you, "Look, we're sorry the Society of Music Lovers did this to you—we would never have permitted it if we had known. But still, they did it, and the violinist now is plugged into you. To unplug you would be to kill him. But never mind, it's only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from YOU."[78]

Thomson then asks [click], *asks you*:

> Is it morally incumbent on you to accede to this situation? No doubt it would be very nice of you if you did, a great kindness. But do you *have* to accede to it? What if it were not nine months, but nine years? Or longer still?[79]

Okay does everyone understand what Thomson is asking here? Whether it would be wrong to unplug yourself from the the violinist, you know, before the nine months are up? Or longer?… I see nodding. Good. Now take a moment to answer, in your head. All right. Recall the plausible-sounding argument? Keep it in mind as I [click] read the next bit:

> What if the director of the hospital says, "Tough luck, I agree, but you've now got to stay in bed, with the violinist plugged into you, for the rest of your life. Because remember this. All persons have a right to life, and violinists are persons. Granted you have a right to decide what happens in and to your body, but a person's right to life outweighs your right to decide what happens in and to your body. So you cannot ever be unplugged from him."[80]

Do you see the parallel? With the anti-abortion argument?… Good. Here's the [click] last bit:

> I imagine you would regard this as outrageous, which suggests that something really is wrong with that plausible-sounding argument I mentioned a moment ago.[81]

This is Thomson's main point. A picture of the parallel [click] might help you see it:

76. Thomson, "A Defense of Abortion," 48.
77. Thomson, 48–49.
78. Thomson, 49.
79. Thomson, 49.
80. Thomson, 49.
81. Thomson, 49.

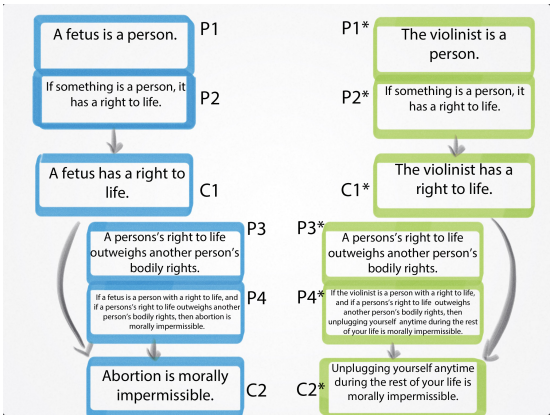

Figure 2.3: Parallel Arguments

> Do you see how the hospital director's argument that you shouldn't unplug *is a lot like* the plausible-sounding one that abortion is wrong? How, for example, in P1*, the director takes the violinist to be a person *just as*, in P1, we assume, for the sake of argument, that a fetus is a person? And, crucially, do you see how, if Thomson is right and you do think *that C2* is false*—that, I mean, it's *not* wrong to unplug yourself—and, also, *if* your thinking this shows that the director's argument fails, then, you see, *then* there's reason to believe the *very similar* anti-abortion argument also fails?… All right, now, what do you think? Is it wrong to unplug yourself? And do you think Thomson's thought experiment succeeds? That is, does it cast doubt on the success of that plausible-sounding anti-abortion argument?

In the ensuing class discussion, as above, a student misunderstands and the teacher corrects:

STUDENT: I've never been hooked up to a famous violinist, and really has anyone, ever?

TEACHER: Well, remember what I told the class, that it's a *thought* experiment, as in an experiment *in thought* or *in the mind*, not out in the world. If what we read were about an experiment in the world, it would make sense to doubt that the hooking up really happened, but it's about an experiment in the mind, and so it doesn't. It doesn't make sense to doubt that what goes on in a thought experiment really happened, or, for that matter, to deny or affirm it, and so on. To do so is to misunderstand the thought experiment. Am I making sense?

STUDENT: So if I uh doubt that the hooking up, or whatever, that goes on in a thought experiment really happened, then I don't get it because "thought experiment" means "experiment *in the mind*"?

This teacher, if successful, removes the student's misunderstanding with—not merely a form of "It's a thought experiment"—but also an analysis of the term "thought experiment." This is because the analysis puts the teacher in position to offer the student a reason not to so respond—e.g., not to doubt that the hook up really took place.

Before I compare this teaching to other forms, I want to bring out how naturally such analyses arise from our grammar. To this end, consider a variation of the last bit of teaching:

TEACHER: What I read, recall that it's a thought experiment—that I called it "a thought experiment." Compare the word "thought experiment" with "experiment." Notice the ah-that-makes-sense click when say to yourself "*thought* experiment" or "A thought experiment is a *thought* experiment."

> OK?… Good. You saw "thought" in "thought experiment" as an adjective, or an adverb, one that modifies "experiment." Now say, "a thought experiment is an experiment in thought."… Good. Notice that it rings true. But not only that. So too "A *thought* experiment isn't a *real* experiment." Also, "A thought experiment takes place in the mind, not in the world." And even "If I read you a *thought* experiment, I didn't read you a description of a *real* experiment." OK?… Here's the point. This also rings true: "If you understand a *thought* experiment, you'll neither affirm nor deny that what's in it *really* happened." Following the very grammar of the term "thought experiment" can get you all the way here. That is, since "thought experiment" means "experiment *in thought*," you misunderstand a thought experiment, such as Thomson's, if you do anything like deny—e.g., affirm, doubt, justify—that what's in it really happened, like you did regarding the violinist hook up.

Now for the comparison. This teaching—like the preceding parent's and professor's above—works neither (i) by *reminding* the student, for whom thought experiments are new, not to misunderstand what goes on in them, nor (ii) by correcting false beliefs about these goings on, nor (iii) by puppy-puddle-prevention-like training. Unlike the preceding parent's teaching, but like the professor's, this teaching does employ a form of the phrase "it's a…" Here, however, the expression "it's a thought experiment" introduces an analysis, a means to clear away the student's misunderstanding, whereas the professor's use of it doesn't introduce any such means. Rather, it alone is to clear away the misunderstanding.

In sum, all three ways of teaching work neither by reminding, correcting, nor training, and all three aim to dislodge a similar sort of misunderstanding. But, to these ends, whereas the parent's with-fable teaching simply directs one *to do* something—that is, has the form

> RESPOND LIKE THIS

—and whereas the teacher's with-analysis instruction *offers a reason*—that is, has the form,

> SINCE THE TERM MEANS THIS, DO NOT RESPOND LIKE THAT

—the professor's with-"It's a thought experiment!" teaching tells the student *not to do* something, *without* giving a reason—that is, has the form

> DO NOT RESPOND LIKE THIS TO THINGS LIKE THAT.

This last use I've tried to clarify with the various preceding comparisons. To draw out its import, I'll now drag this abstractly characterized use of "It's a thought experiment!" back down to earth.

The Import

In this abstract characterization—i.e., "do not respond like this to things like that"—there are two models, or objects of comparison, namely, that which the "this" and the "that" point out, that is, respectively, the denying, or affirming, and Black's thought experiment. More specifically, just as we use cookbook diagrams to guide our pastry-rolling—"we roll it just like this"—the above professor who says, "It's a thought experiment!" uses Black's thought experiment to guide the student's responses to similar things; or, alternately, just as we model pie crusts on pastry pictures—"the crust should look just like this"—the professor models, upon this thought experiment, *what* the student should not do in similar circumstances. And the same goes for the student's affirmation or denial; the professor uses it to guide the student's responses to what's like Black's thought

experiment—or, alternately, she models, on this response, *what* the student should not do in similar circumstances.

To be sure, I'm not saying, first, that the student who learns must go on to *compare* this thought experiment or that affirming, or denying, to anything else; it may be like a cookbook diagram not needing to come to mind to roll rightly once you've learned the rolling technique. The student need not, for example, write down a rule, such as "If you hear a passage like this one (Black's), don't do anything like denying that the spheres exist," then memorize it, and, finally, recall those words and follow the rule expressed upon encountering similar passages. Instead, the student may repeat such words, just once, and then, without recalling them, follow the rule, i.e., stop responding in the inept way; or might feel slightly uneasy, and, when encountering another thought experiment, feel that unease again and not respond ineptly; or may go, "hmm," and simply stop responding so; or, without writing down or repeating the rule, or feeling or doing anything special, may simply stop responding ineptly. Second, I am also not saying that the teaching may be successful only if the professor intends it, e.g., *plans ahead of time* to use these models, objects of comparison. Third, I'm also not, of course, saying that a student, in a dialogue with different objects of comparison, couldn't learn more or less the same thing—e.g., that the one learning in light of Black's thought experiment and a denial couldn't learn more or less the same thing as another in light of Thomson's and an affirmation.

Let us continue our Earthward pulling. What is learned? That is, what is it not to respond like this to things like that? What the student learns stretches out, as it were, along two lines. One moves past ceasing *to affirm or deny* propositions in Black's thought experiment's story-like part, the other past *that thought experiment.* To illustrate the first, the student also *won't* do such things as secretly think that the propositions are true; or that they're false; or wonder whether only the spheres exist; or exclaim in amazement, after hearing Black's thought experiment, "Wow, I can't believe that's all there is!"; or even, perhaps, joke, "So Black's a dualist; for him, only two substances exist." To illustrate the second, past ceasing to affirm or deny propositions in "the story" of *what's like Black's thought experiment*, the student also will *not* deny, e.g., if responding to *Thomson's thought experiment*, that anyone was ever kidnapped to save a famous violinist. Along this line, there is also, among other things, not doing so in response to what is called "a thought experiment," or the like, or what has a name with the form "So-and-so's Such-and-such Thought Experiment," or what begins with a word like "suppose," which introduces something like those spheres, which in turn leads to something like Black's metaphysical point. To be sure, what the student learns may also extend along both lines, e.g., when he or she won't *declare* in response to *Thomson's thought experiment* that we must never again let the Society of Music Lovers engage in a violinist-saving-but-rights-violating kidnapping. To sum up with an analogy, what the student who learns wouldn't do extends the way this argument does: Red roses are lovely; Whatever is like a red rose is also lovely; So *white* roses are lovely, and red *tulips* are lovely, and *white tulips* are lovely, and so on.

This kind of rule we sometimes overlook when thinking about what thought experiments are. For example, to explain them, we often appeal to imaginings, as discussed above, without any thought to simply stopping the sorts of responses specified by the rule; and, an appeal instead to such a rule might explain equally well in some but not all cases. This overlooking, then, may partly explain why it seems so implausible to deny, as I do above, that, in short, imaginings are common to thought experiments.

Five Sources of Doubt

To end the chapter, consider five sources of doubt.

First, our student above could say, "but there are no such spheres," and yet still understand Black's thought experiment, were he to mean that there aren't any in the genuine article. To elaborate, such a student may go

on to say: "The spheres are not made of iron. In the real thought experiment, the one in Black's original journal submission, they're made of iridium. Apparently, the editor 'corrected it'." This would be to make an interpretive point. By analogy, you could deny the fox's existence but still understand *The Fox and the Mask* fable were you to argue for an interpretation as follows: "I read that, in the medieval Latin tradition, this story isn't told of a fox but of a wolf. The story in that tradition is the authentic one. So, since there's no fox in the fable, no fox says anything in it."[82] Failing to recognize interpretive points as such, the denial may resemble a criticism of a report. This, in turn, may give rise to a misplaced doubt about my claim that the denial is confused.

Second, to arrive at a similar misplaced doubt, we may, if not distinguishing imaginings from assumptions about them, think that the student can, without misunderstanding the thought experiment, deny that the imagined spheres exist. Recall the student who disagrees with Black's point. This student says that the two spheres are in different places and so do not have all the same properties. The professor draws out the idea, putting it in terms of Black not describing what he takes himself to be describing—i.e., two spheres with identical properties—but instead describing two spheres with at least this difference, that they're in different places, i.e., have different spatial properties. In this light, we see that the student disagrees with an assumption as opposed to something asserted in what Black called his "description." Now, this disagreement may be mistaken for that of the other student who denies that the world contains no identical spheres. If so, then, since the first student understands Black's thought experiment, it may appear that one can also understand it and make the second student's denial. This, in turn, may give rise to the misplaced doubt.

But how do such assumptions differ from the imaginings they're about? The imaginings "come first." Here is the germ of how I conceive the difference. Normally, learning not to deny the spheres' existence, or the like, in response to Black's thought experiment, and what's like it, *comes before* learning to disagree with that thought experiment's assumptions, e.g., that the spheres have identical spatial properties. For example, normally, a student learns not to make anything like such in-response denials—then learns to disagree with those assumptions—and, were a student not to learn to not make the denials, that student would not learn to disagree with the assumptions.

Third, how could the student possibly have learnt anything from hearing "It's a thought experiment," given that he does not yet know what the word means? He can do so as we saw in the above language games. But one may be unconvinced and argue that he cannot, as follows. For any sentence, if it has a meaning, that meaning consists in its truth conditions. Therefore, to explain what a sentence means you must explain the conditions under which it's true. Thus, to explain what the sentence "It's a thought experiment" means one must get across that the bearer of "it" falls under the predicate "is a thought experiment," provided that the one being explained to knows what this singular term and predicate mean. But, above, *ex hypothesis* the student doesn't know it and, *a forteriori*, doesn't understand the predicate "is a thought experiment." Hence, the student cannot understand "it's a thought experiment." Hence, since one cannot learn anything from a sentence one doesn't understand, the student can't learn anything from it, contrary to what I claim. This argument, however, fails, at least insofar as it depends on a bad analysis of the expression "It's a thought experiment"—specifically, one that ignores its use. To see this, compare the objector to someone who takes a loud, snarling "Go to Hell!" to mean "I order you to move yourself from your present location to another named 'Hell'," and reasons that, since he can't tell whether the angry person meant the hellish underside of heaven or heavenly Hel, Poland, he hasn't been told where to go.

Fourth, don't I confuse thought experiments with accounts of them? For example, you might take

82. Cf. Gibbs, *Aesop's Fables*, Fable 550.

expressions of mine, such as "the professor read a student Black's thought experiment," and say they're nonsense—like "I read a half-marathon." By contrast, "I read *an account of* a half-marathon" makes sense. So too "the professor read a student *an account of* Black's thought experiment." Alternately, one can say, "I *ran* a half-marathon." So too "the professor *carried out* Black's thought experiment" or "the student *performed* it." Doesn't the confusion infect my comparisons between language games and consequently give rise to doubts about the insights I try to draw from them?

In reply, notice that I can say "she read you *The Fox and the Mask*." So too "she read you a recipe" and "she read you a plan." I can also say "she performed *The Fox and the Mask*, e.g., wore a fox suit, held up a mask, etc." So too "she followed a recipe" and "she carried out a plan." Moreover, "she read *The Fox and the Mask*" doesn't mean "she read a description of *The Fox and the Mask*." For one could be true without the other being so. So too "she read a description of a recipe" and "she read an account of the plan." Now we can model our use of "thought experiment" on that of "fable." That is, just as we can say "she read you this fable," we can say "she read you this thought experiment," and just as we can felicitously say "she performed this fable, e.g., acted it out," we can say "she performed this thought experiment, e.g., acted it out," and just as we can say "she described this fable, e.g., summarized it," we can say "she described this thought experiment, e.g., summarized it." So modelled, i.e., taking "fable" as a certain sort of object of comparison, it's not nonsense to say "she read Black's thought experiment." To be sure, modelling it on "fable" doesn't work for "she carried out a thought experiment," since it doesn't make sense to say "she carried out a fable." But, at least for all that's been said, we've no reason to take modelling it on "experiment," from which the carrying-out language presumably derives, to take precedence.

Finally, fifth, about the stop-sign-like use of the term "thought experiment," how could it possibly bear on the term's calling use? The worry is that, if it doesn't, there may be no relevant connection between it and our unreflective concept of a thought experiment.

Suppose that Young Hal hasn't yet seen a hammer or heard "hammer." To teach him about hammers, you show him one, nail a nail with it and say, "it's a hammer." He then calls the next hammer-shaped, nail-nailing thing he sees "a hammer." Also, sometimes, when told, "hammer," he brings you a hammer or uses one to nail a nail. But, other times, neither does he call a sledge hammer at work "a hammer" nor does he bring such a hammer when he hears, "hammer." And so on. Here is what he learnt: how to nail nails with some hammers, what shape some hammers have, and, in some cases, what's called a "hammer." And here is what you used to teach it: a hammer, a use of it, and the phrase "it's a hammer." It's as if you told him, "Call something shaped like *this* and used like *this* a 'hammer'." Notice that your use of "hammer" here is clear as it stands, no need for an account of how Hal's cochleae work. Similarly, you might use "It's a thought experiment!" not only to tell someone, as it were, "do not do anything like *this* in response to anything like *that*," but also, "call a 'thought experiment' that which is like *this* and, in response to which, you're not to do anything like *that*." In ways like this, the rule bears on calling uses and so on the concept's correct application.

3 On Literature as Thought Experiment

Are novels experiments? Some are, thinks Zola, namely, ones written by "experimenting novelists." These novelists try to reveal certain mechanisms, specifically, those behind social and individual phenomena. Also, trying to do so, they stick to scientifically established facts and laws, except when writing about what hasn't been so established, in which case they stick, so far as possible, to observation and experiment.[1] His main example of an experimental novel is Balzac's *La Cousine Bette*. To write it, says Zola, Balzac observed the ravages of amorous men, and then, relying on his experience and never departing from natural laws, he experimented. That is, he put a fictional character, the amorous Hulot, through various trials, which revealed the mechanism of amorous passion, yielding scientific knowledge. Then, experiment complete, he repeats it; that is, novel written, the public reads it.[2]

Are novels thought experiments? Zola doesn't say, but Mach does. He says they are, since novelists experiment with thoughts—i.e., imagine conditions, see what they expect to happen, and surmise what will.[3] To illustrate, let us recast Zola's example. Balzac imagined amorous trials through which to put Hulot, expected certain results, and then worked them out in thought. Mach would then, perhaps with Zola's approval, label experimental novels like Balzac's "thought experiments," distinguishing them, as he does all novels, from experiments in general; however, he would deny that any experimental novels exist, since, for him, and much to the would-be consternation of Zola, no novels yield scientific knowledge. That is, he thinks novelists, unlike good thought experimenters, neither imagine wholly realistic conditions nor actual consequences thereof, and, thereby, they let their thinking stray from reality. But that's not all. To the consternation of both, Roy Sorensen, in effect, denies novels both the "experiment" label and the "thought experiment" one too. For him, the complexity of aims in, say, Balzac's novel dilutes it down past the point of experiment, as we'll see in §3.4.2.

Why care whether or not novels are thought experiments? If they are, we can explain fictional learning. That is, if asked how we could possibly learn from works of literary fiction, such as the novels which populate the Western Canon, we could answer that they are thought experiments. Like Zola, some contemporary philosophers, such as Catherine Elgin and David Davies, give explanations along these lines. My central contention is that, in so doing, Elgin and Davies risk turning from something we want explained—i.e., ordinary literary learning. To argue for it, I won't of course deny, as Mach might, that we learn from works of literary fiction that are thought experiments. Rather, like Sorensen, I argue, ultimately, from differences between the two. They concern complexity but also imaginings and interpretive freedom.

This chapter has four parts. In §3.1, I specify my central contention. In §3.2, I argue for it. In §3.3 and §3.4, to develop this argument, I explain the differences.

1. Zola, *Le Roman Expérimental*, 52.
2. Zola, 8.
3. Mach, "On Thought Experiments," 136.

3.1 Contention Specification

To specify my contention, first, I specify a problem about literary learning. Then, for this problem, I sketch a solution schema. This in hand, I characterize two solutions—namely, Davies' and Elgin's. That done, I'm in position, in §3.1.1, to make the contention specification.

Throughout, I explain by making comparisons to the position of one Edward Davenport. Its troubles help to shed light on those which are my focus.

Problem Specification

To clarify our problem about literary learning, consider how Davenport tries to raise one like it. He does so as follows. On a traditional view of art and science, they're "polar opposites." That is, art unlike science expands our "emotional and noncognitive… awareness or sensibility"; and, science unlike art advances our "cognitive knowledge."[4] If we take this view, we seem unable to learn from literature about the world beyond the text.[5] But readers report such learning and, without it, we cannot understand or appreciate literature.[6] So, we have a problem: traditional view or literary learning?

This problem, read without charity, isn't well-formed. We obviously learn from literature, and that "traditional view," with its absurdly simple division between art and science, is hardly plausible. Read with charity, however, matters stand differently.

Before we get there, to convince yourself that we obviously do learn from literature, recall three cases. First, when reading studiously, we're often led to look up words, and so, when not merely reminding ourselves, we learn from dictionaries or encyclopaedias; in this way, we learn about the world from the work in virtue of what we must do to read it well. Certain interpretive styles, especially "New Criticism," may prohibit doing so, but we often do it anyway. Second, from reading a novel, we often learn about its origins. We learn, for example, that Tolstoy wrote *Anna Karenina* or that Richard Pevear and Larissa Volonkhonsky translated it. Alternately, from Jack Kerouac's *On the Road*, commonly read as a roman-à-clef, we learn from the travels of its protagonist, Sal Paradise, that its author travelled across the United States.[7] Third, reference materials and origins aside, we often learn about the world from a work's setting. That is, some phrases help to fashion a story's setting, and, while reading, we often believe the truths that some of these phrases express. Some such truths are particular. We read, for example, in every one of William Gaddis' novels, many sentences without distinguishing setting from historical description, and, if what we read is true, we learn, in *The Recognitions*, that participants at the First Council of Nicaea debated whether Jesus and God are the same substance or merely like ones;[8] in *JR*, that Bizet died three months after *Carmen* was produced;[9] in *Carpenter's Gothic*, that Masai warriors believe "that all the cattle in the world belong to them"; [10] in *A Frolic of His Own*, that the Battle of Antietam was "the bloodiest single day of the entire [American] Civil War";[11] and in *Agapē Agape*, that "Vaucanson's loom for figured silks" prefigured the player piano.[12] Conversely, some such truths are quite general. For example, we finish reading whole novels, and, in light of some fairly common background knowledge as well as certain passages in the work itself, we feel familiar with a swath of the world—from *War*

4. Davenport, "Literature as Thought Experiment (On Aiding and Abetting the Muse)," 279.
5. Davenport, 279.
6. Davenport, 280.
7. Kerouac, *On the Road*.
8. Gaddis, *The Recognitions*, 9.
9. Gaddis, *JR*, 117.
10. Gaddis, *Carpenter's Gothic*, 121.
11. Gaddis, *A Frolic of his Own*.
12. Gaddis, *Agapē Agape*, 21.

and Peace, of the 1812 French invasion of Russia and its precursors, in light of some historical knowledge and Tolstoy's straightforward arguments against "great men in historical events";[13] from Melville's *Moby Dick*, of whales and historical whaling, in light of some nautical knowledge and the (in)famous cetology taxonomy chapter;[14] and, from Proust's *In Search of Lost Time*, in light of historical knowledge and straightforward explanation, of life in France around the turn of the 20th century, e.g., the Dreyfus Affair, the aristocracy's decline, homosexuality, art, etc.[15]

Now, turn to a charitable reading of Davenport's problem. If read in light of his main example of learning about the world from literature, it's not at all obvious that we do so learn, and in this respect the problem is well-formed. Let me explain. This main example comes from the plot of Eliot's *Middlemarch* and not from its setting alone, or its publication information, or the reference material we use to read it well, and so on. More specifically, it comes from the story of young Dorothea's unhappy marriage to old Casaubon. The marital drama in this story teaches us, he thinks, among other things, that—against the traditional idea that sexual problems explain the failure of marriages between young and old—"marriage is not based on sex, but is based on love and mutual respect."[16] In light of this example, we might charitably read the problem as follows. We seem to gain science-like knowledge from stories in literary works, but we cannot so learn from art, since it and science, on a traditional view, are polar opposites.

Even on this charitable reading, however, the problem isn't well-formed, since this polarity view remains implausible. To better formulate it, we will, instead of trying for a more charitable reading, simply replace the offending view. Specifically, we'll replace it with one in the same spirit that isn't obviously implausible. An admirably simple candidate is due to Catherine Elgin: "works of fiction neither are nor purport to be literally true."[17] We'll pass this over because, in light of the above cases of learning, it needs qualification, such as to substitute "works" for "stories in works." A less admirably simple candidate, but one without this qualification issue, is due to David Davies: "at least as popularly construed—fictional narratives are (by definition) fictitious, or, at least, written without the concern with mapping reality that is supposed to guide non-fictional narratives."[18] Adopting this candidate—and understanding "true" and "false" in terms of the concern in it—the problem becomes

> How could we possibly learn about the world from stories in works of literary fiction, given that they're either false or, if true, not written to be so?

In what follows, we focus on this problem.

Solution Schema

Davenport's solution to *his* problem—i.e., literary learning or traditional view?—consists in replacing the view. In particular, he aims to replace it with one on which literature is in its methods partly rational and in its teachings partly cognitive.[19] For him, one such method so teaches, and it consists in certain sorts of writing and reading—which he calls "thought experiment."[20] This teaching method, in effect, aims to solve *our*

13. See, especially, the Epilogue, Tolstoy, *War and Peace*, 1215–1225, and IV-2-8, Tolstoy, 1077–1078. For the quotation, used merely to express the idea, see the Appendix, Tolstoy, 1314.
14. Melville, *Moby Dick or the Whale*, 119–131.
15. Proust, *À la Recherche du Temps Perdu* An example of straightforward explanation, that of "inverts," follows the opening scene in Volume Four, *Sodome et Gomorrhe*.
16. Davenport, "Literature as Thought Experiment (On Aiding and Abetting the Muse)," 301.
17. Elgin, "The Laboratory of the Mind," 41.
18. Davies, "Thought Experiments and Fictional Narratives," 31.
19. Davenport, "Literature as Thought Experiment (On Aiding and Abetting the Muse)," 279.
20. Davenport, 302.

problem—i.e., the one about learning from stories in works of literary fiction. Responses of this general kind will be our focus.

To characterize this kind, I'll contrast it with three others.[21] The first two don't solve the problem; they reject it. One denies that we learn about the world from stories in literary fiction. It would be as if we read every work as pure fantasy, as you might read, say, *The Lord of the Rings*. The other denies that the stories are neither true nor meant to be, and it affirms that we learn from them as we do true reports. It would be as if we interpreted every work as veiled autobiography, as you might read, say, a story about the fictional character Gately in *Infinite Jest* as about the real-world Craig.[22] The other two responses don't reject the problem; they solve it. One is that we learn directly from what we read. For example, take John Gibson's solution. Very roughly, it is that, reading the stories, which need not be true and aren't meant to be, we learn about certain real-world standards of representation. By analogy, examining the toys by which children learn to tell trains from airplanes, we can learn about how for them such vehicles differ in essentials. This learning is direct—i.e., builds no bridge from the story to the world we learn about—because, roughly, the standards we learn are the stories themselves.[23] For example, on this account, reading a certain story in *Alice in Wonderland*, we might learn what it is to go "down the rabbit hole," insofar as the story itself is, or is part of, our standard for applying that popular expression. Finally, the other kind of response, that which is our focus, is a solution but not a direct form of learning. Rather, it is that we learn and that we do so in virtue of certain ways we think about the story. Put another way, our thinking, as it were, builds a bridge from the fictional to the real world, making learning possible.[24] Davenport's solution, for instance, is of this kind. For him, we don't learn about the real-world bases of marriage simply reading the story of Dorothea's marriage in *Middlemarch*, as we would were the story a standard for representations of good marriage. Rather, for him, we do so in virtue of how we think about what we read and, specifically, by means of the thought experiment we perform.

As I said, this last is the general kind of response on which we'll focus. The more specific kind we'll focus on closely resembles Davenport's solution. Here is a schema for it:

> When we learn about the world from stories we read in works of literary fiction—even though they're false or at least not meant to be true—we do so in virtue of thought experiments in our thinking about those works.

Two Solutions

My main claim concerns two solutions of this specific kind. To introduce them, consider a third, based on Davenport's solution, and two gaps in it.

To give this introductory solution, we fill in the schema, as follows. While appreciating a literary work, we make certain comparisons and, thereby, test "the plausibility of ideas" about society, which the author dramatized.[25] To elaborate, in our literary appreciation lies "the method of reading" or "secondary thought experiment." Following this method, readers compare ideas the author dramatized against their "knowledge and experience," thereby testing the ideas for plausibility. This method, moreover, comes after "the method of composition" or "primary thought experiment." Following this method, an author discovers which way of telling a story dramatizes a plausible idea, which idea readers will "independently" check.[26] It is this checking,

21. Here, in large part, I follow Gibson, *Fiction and the Weave of Life*, 109–110.
22. Cf. Wallace, *Infinite Jest*, 55, 476, 902. Also: "Friends closed elevator doors on his [Craig's] head for fun when he was a teenager, a detail [David Foster] Wallace would put into *Infinite Jest*" (Max, "Every Ghost Story is a Love Story: A Life of David Foster Wallace," 141).
23. Gibson, *Fiction and the Weave of Life*, 123.
24. Gibson, 110–111.
25. Davenport, "Literature as Thought Experiment (On Aiding and Abetting the Muse)," 301.
26. Davenport, 302.

in our thinking about so-written stories in works of literary fiction, that is the thought experiment in virtue of which we readers learn about the world—even though the stories are false or else not meant to be true.

Now consider the two gaps. First, this solution doesn't explain how what's learnt counts, in a significant way, as partly scientific. After all, on it, we come to know, at best, what is true to life, i.e., what coheres with our, perhaps limited, knowledge and experience. To bring this point out, add to this solution that which Zola builds into his experimental novels, namely, that the knowledge against which we test is, with certain exceptions, of scientifically established facts and laws. In that case, the "plausible ideas" we arrive at would be less likely to stray from reality, as Mach would have it. Without this addition, however, the solution's learning may easily so stray. So it doesn't count, in a significant way, as partly scientific. Second, connectedly, it doesn't explain empirical learning. To bring out the point, this learning, in light of Davenport's main example, concerns the world, e.g., the bases of marriage. So the solution should address a question like Thomas Kuhn's well-known one: "How… relying exclusively upon familiar data, can a thought experiment lead to new knowledge or to new understanding of nature?"[27] Yet no such question is addressed.

In light of these gaps, to solve our problem, we might follow Zola and appeal to a scientific theory. That is, we might, first, find a scientific explanation of how thought experiments work. Then we might apply it to thought experiments in our thinking about stories in works of literary fiction. Finally, we might point out that, thereby, we've explained how one learns about the world from such stories, even though they're false or at least not meant to be true. We might also do all this—that is, apart from filling the above gaps—because extending such a theory's application increases its explanatory power.

Take, for example, Nancy Nersessian's[28] and Nenad Miščević's[29] theories of thought experiments. On these theories, by manipulating mental models, we can gain access to tacit empirical information and, ultimately, reach new knowledge of the world. David Davies thinks we can apply such a theory to solve our problem. That is, he argues—and this is the first of the two solutions on which I'll focus—that it explains our learning from certain stories in works of fiction, namely, those which are thought experiments.[30] For example, on his view, when we appreciate a Henry James novel, we manipulate certain mental models, gain access to tacit empirical information, and come to know about certain complex human relations.[31]

Now, to introduce the second of the two solutions I'll focus on, we may not want to apply, as Davies does, a theory *of* thought experiments to stories in literary fiction. We may, instead, want to apply a theory to such stories *in light of* its application to thought experiments. Thereby, we could be agnostic about whether any of these stories are thought experiments and hold merely that we can *regard* some of them as such. To clarify the difference, consider how Catherine Elgin's early account differs from her later one. On the early one, to explain how we come to a new understanding of the world, she argues that thought experiments teach us concepts and that these concepts help us to better organize our worldly experience.[32] Then she argues that the same goes for some works of fiction, since they are thought experiments. For instance, from Big Brother's systematic gerrymandering of historical records in *1984*, she thinks that we gain a concept which helps us see that intersubjective agreement doesn't suffice for justification.[33] On Elgin's later view, by contrast—and this is the second of the two solutions I'll focus on—she applies a theory, which posits an independently described mechanism, i.e., exemplification.[34] That is, she posits, as it were, samples by which we can understand that in

27. Kuhn, "A Function for Thought Experiments," 241.
28. Nersessian, "In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling."
29. Miščević, "Mental Models and Thought Experiments."
30. Davies, "Thought Experiments and Fictional Narratives," 43–44.
31. Davies, "Learning Through Fictional Narratives in Art and Science," 65.
32. Elgin, "The Laboratory of the Mind," 47–48.
33. Elgin, 50-51.
34. Elgin, "Fiction as Thought Experiment."

the world which the samples are of. Thereby, she explains learning from experiments, from thought experiments, and—crucially, in light of these explanations—from stories in works of literary fiction. This explanation doesn't require that any stories *be* thought experiments. Still, as I'll explain below, it has us regard the ones as the others. That is, it limits our purview of the stories to features shared by experiments and thought experiments.

In sum, I'll focus on Davies' and Elgin's solutions. More specifically, my primary concern will be the identity claim and "regarding as" assumption on which they depend, respectively. Other contemporary accounts I'll consider, apart from Davenport's, include those by Noël Carroll and Roy Sorensen.

3.1.1 Central Contention

My central contention is that to give either kind of solution, insofar as it depends on identity claim or regarding-as assumption, is to risk losing one's grip on literature ordinarily read. To clarify, I'll give an analogy and restate the contention in light of it. In particular, before the restatement, I'll explain a similar contention, namely, that Davenport, making his identity claim, risks losing his grip on it. To do so, I'll explain a form of literary appreciation, his identity claim, a way these two diverge, and why, because the two do so, he risks losing his grip.

First, recall some comparisons you've made, while ordinarily trying to understand a work of literary fiction, between story and world. Can we call such comparisons "testing the plausibility of an idea"? Some we cannot. For instance, we cannot call the following "testing": a comparison, in Gaddis' *JR*, between the fictional Myrna and the real Joan Bennett when she dyed her hair black—insofar as it merely develops that character. But others we can. For example, I can recall myself comparing events in my life with those I was reading about in Proust's *The Captive*, when jealousy repeatedly inflames the narrator's love for Albertine. Specifically, reading about such jealousy, and feeling as though my feet were off the ground, I tried to recall a case of it in my own life, i.e., a case like that which I was reading about—and, once successful, I was able to get my feet down and into the narrator's shoes. We can call this comparing "testing the plausibility of an idea." To be sure, consider two ways it resembles measuring to test a hypothesis. One is that I used a recollection, like a ruler, to see, as if by measuring, whether or not it's realistic that the narrator's jealousy is inflaming his love. The other is that, finding the inflaming lifelike, I "identified," or empathized, with the narrator, as if, finding that a hypothesis agrees with my measurements, I confirm it.

This form of literary appreciation now described, consider the identity claim Davenport makes. He gives two arguments for it. First, it's the conclusion he draws from his interpretation of Dorothea's marriage story in *Middlemarch*: "it is part of the nature or structure of a story that we can use it to test the plausibility of ideas."[35] That is, stories are, among other things, for testing the plausibility of ideas—"for," that is, as a hammer is for hammering but not as it is for juggling. To be sure, we can understand his claim, and especially the expression, "nature or structure," otherwise, i.e., read it such that he claims merely that stories *might be* used to test an idea's plausibility—"might be," that is, as a hammer might be used to hammer or else to juggle; however, we shouldn't read it so, for two reasons. This testing, first, is, and, second, should be, for him, a means to "understand and appreciate" works of literary fiction—i.e., a proper use, as is hammering of a hammer—and not merely a means for any old use of these works. That for him testing *is* such a means, we see it explicitly stated in the other argument he gives for this identity claim. It is that, since we test the author's ideas to "understand and appreciate" fictional stories, they're experiments, and since it occurs in the imagination,

35. Davenport, "Literature as Thought Experiment (On Aiding and Abetting the Muse)," 301.

they're thought experiments.[36] Furthermore, from this argument it follows that, for him, testing not only is but *should* be such a means, since, in the argument, it's by means of testing that we understand and appreciate the stories and not merely that, in some way or other, we use them. In sum, the key point here is that he claims that stories in works of literary fiction are thought experiments. That is, as we use hammers for hammering, we use them for a kind of literary understanding and appreciation, specifically, for testing the plausibility of ideas they dramatize about the social world.

Now, the use of comparisons in this identity claim diverges from that in above appreciation. That is, in line with that in the identity claim, as I said, to understand Proust's story I compared it against my knowledge and experience, and, this comparing, I can call it "testing the plausibility of an idea." But, at odds with the identity claim, I can't call this test "one for learning about the world beyond the text." To be sure, suppose that the test were for such learning, specifically, for testing the plausibility of the idea that, in the world, jealousy inflames love. In that case, I would have failed to understand and appreciate the story, since the test would have failed. That is, I'd have failed, since I'd have been making such blunders as (i) cherry picking evidence, given that I merely tried to recall one "confirming" example and exactly zero "disconfirming" ones, and as (ii) hastily generalizing, given that I recalled only my experience and then only the first one that came to mind. But I did understand and appreciate it, insofar as I was able to empathize with the narrator.

Finally, this diverging use of comparisons raises a questions about the identity claim's scope. That is, in light of the divergence, we can ask: Do we, as a matter of course in our everyday reading, thought experiment when understanding and appreciating stories in works of literary fiction? And we can give a qualified answer. Perhaps not, since our doing so may, unlike the Proust case, not be characteristic of our everyday reading. Put another way, Davenport got hold of a way to interpret, but, for all he says, this way is not a characteristic everyday one, that is, one we in our everyday lives follow as a matter of course when understanding and appreciating stories in works of literary fiction. That is, he, as it were, risks losing his grip on how we normally read them. Thereby, he risks missing, in his explanation of how we learn from them, how we ordinarily do so.

Likewise, my central contention is that to explain, as either Davies or Elgin does, is to risk losing our grip on something we want explained. Specifically,

> If, to explain how we could possibly learn about the world from certain stories in works of literary fiction, given that they're either false or, if true, not written to be so, proceed, as Davies does, by means of claiming these stories are thought experiments, or, as Elgin does, by regarding the ones as the others, then, we might explain uses of stories taken from literary works but, in so doing, inadvertently miss characteristic everyday forms of literary appreciation.

About this contention, three points. First, it borrows much. It borrows a strategy from a well-known argument due to Peter Lamarque and Stein Haugom Olsen, namely, in short, that, since literature has themes but no theses, it's not about the world—which undermines explanations of how it teaches truths about the world.[37] Also, in light of Peter Kivy's well-known response to that argument, I focus on everyday reading.[38] Finally, I draw on Sorensen's idea that we can read a novel as a thought experiment but deny that it is one.[39]

Second, the contention matters. This is so even though it raises difficulties for solving a philosophical problem without advancing a better theory. For it's a step toward clearing the problem away instead of advancing a theory to solve it. That is, if true, we can use it to give an account somewhat like Lamarque and

36. Davenport, "Literature as Thought Experiment (On Aiding and Abetting the Muse)," 301.
37. Lamarque and Olsen, *Truth, Fiction, and Literature: A Philosophical Perspective*, 321–338.
38. Kivy, *Philosophies of Arts: An Essay in Differences*, ch. 5.
39. Sorensen, *Thought Experiments*, 222–3.

Olsen's as well as the above two reject-the-problem ones—i.e., one on which certain misunderstandings give rise to philosophical problems about how we learn from stories in works of literary fiction. After all, the contention concerns a failure of two prominent theories to so much as get a sure grip on what's to be explained.

Third, finally, the contention coheres. That is, it hangs together with earlier chapters, in three ways. First, it aims at a problem solving account, as we just saw, one that is both in line with the approach extrapolated in §1.2.1 and similar to two accounts given last chapter. One of the two, from §2.1—to help clear away a problem about explaining what we know thought experiments to be—appeals to a misunderstanding about definition. The other, from §2.3—to help clear away the same problem—explains why imaginings seem essential to thought experiments. To be sure, the contention itself doesn't amount to such an account, but it helps to give one. That is, it's a step in that direction, much like pointing out cases in which we obviously learn from literature, as we did above, removes part of the problem. That was the first way the contention coheres with earlier chapters. Here is the second. To argue for the contention, in line with §1.2.2, I will neither advance a theory, dig for an essence, nor interfere with actual usage. Third, this argument, although it makes no use of language game comparisons or a family resemblance account, which is at odds with §1.1.1 and §1.1.2, it nevertheless embodies the spirit of these sections. That is, my argument often focuses on actual usage, e.g., on everyday literary practice. Also, it never assumes that every concept has an essence, e.g., when examining how literary stories and thought experiments differ.

3.2 Main Argument

If certain stories are thought experiments, a theory of the latter can apply to the former. One route to justifying this applicability is to argue for the identity claim. Various philosophers do so.[40] Some such arguments resemble Mach's and Davenport's, mentioned above. They are that the relevant stories satisfy a certain criterion for calling something a "thought experiment." Here are two issues that arise for these arguments. First, in light of last chapter, the criterion threatens to mislead us, since it often depends on a picture of thought experiments as experiments in thought. Second, and more important here, giving the arguments, in light of three weakening disanalogies explained below, we, like Davenport above, risk losing our grip on what we want explained. This brings us to Davies' route.

His route goes around these issues, since he argues for identity without a criterion.[41] His argument, at a glance, runs as follows: thought experiments are narrative fictions on a "make-believe theory of fiction"—so, some works of narrative fiction are thought experiments. This roundabout route, however, doesn't get any further, since, as we'll see, two difficulties undercut his argument: first, thought experiments aren't narrative fictions on his theory of them; second, the argument, if deductive, has the wrong conclusion, and, if inductive, faces the same three weakening disanalogies. The result will be that Davies' route, no less than Davenport's above, ends up risking our grip on the stories as literature ordinarily read.

Another alternate route is Elgin's later one.[42] It doesn't have us apply a theory *of thought experiments*. Rather, we're to apply a general theory, of exemplification, to them and, in light of this application, apply it to the stories. The upshot is that we need no identity claim, criteria, or definition, or arguments for any of them. All we need to do is regard the stories as we do thought experiments when we come to apply a theory to them.

40. Elgin, "The Laboratory of the Mind," 47–48, Carroll, "The Wheel of Virtue: Art, Literature, and Moral Knowledge," 7–10, and Davies, "Thought Experiments and Fictional Narratives," 31–33 or Davies, "Learning Through Fictional Narratives in Art and Science," 52.

41. Davies, "Thought Experiments and Fictional Narratives," 31–33, Davies, "Learning Through Fictional Narratives in Art and Science," 52.

42. Elgin, "Fiction as Thought Experiment" but not Elgin, "The Laboratory of the Mind."

But again this doesn't get us any further. The difficulty I find with this approach arises, again, from those same three disanalogies. It is that, if the approach ignores the three and has us so regard certain stories, we, again like Davenport, risk losing our grip on them as literature ordinarily read.

To sum up, here is the main argument in a nutshell. Either by trouble applying a theory of fiction or, ultimately, by ignoring certain disanalogies between thought experiments and the stories, one who, to apply a theory to the stories, relies on an identity claim, like Davies, or else on "regarding as," like Elgin, risks losing one's grip on the stories as literature ordinarily read.

Before further explaining these difficulties for routes like Davies' and Elgin's, let me assuage three worries about my concern with literature ordinarily read, i.e., with everyday appreciation of stories in works of literary fiction.

Three Worries about Literature Ordinarily Read

Here is the first. Why should we want to explain learning in such appreciation as opposed to that in special cases? Well, when we ask our question about learning from stories in works of literary fiction, we want to know about the phenomena as they are. We do not want, or else want merely secondarily, to know about possible or new phenomena invented answering the question. To be sure, an explanation of special cases may tell us about everyday ones. But it may just as well not. After all, a one-sided diet of special-case answers may well skew our understanding of the phenomena.[43]

But errors surely infest everyday appreciation, and so why should we want to explain everyday appreciation as opposed to that of literary critics, i.e, the experts? First, everyday appreciation, errors and all, is itself interesting. Second, it gives accounts like Elgin's and Davies' a better run for their money. That is, whether we advance an account which has us regard some of the stories as thought experiments or which takes them to be identical, if we focus on the critics, we squarely face Lamarque and Olsen's well-known argument that, in short, since the critics deal in themes but not theses about the world, literary works have no such theses for us to learn.[44] Were such an argument to go through, those accounts wouldn't in fact be about literary works. To attribute a thesis about the detriments of certain political and socio-economic systems to *1984*, for example, would be a mistake—i.e., to turn away from that literary work. If, instead, we focus on everyday literary readers, we can more easily leave critics aside and so get around the argument. Indeed, replies to this argument such as Peter Kivy's, which rely on an account of ordinary reflection about works after and between readings, i.e., on their "reflective afterlife," would become stronger.[45]

***

Here is the second worry. Don't I owe some characterization of everyday literary appreciation? Perhaps. To give one, I will: first, criticize an analysis of literature in light of such appreciation; second, point out the examples I'll focus on; third, explain how "everyday appreciation" is coherent; and, fourth, sketch a swath of such appreciation.

Consider an analysis Lamarque argues for, namely, his third sense of "literature": "fine writing of an imaginative/creative kind imbued with moral seriousness."[46] In some cases, the analysis works well. We would do well using it to shelve novels in a bookstore's Literary Fiction section or, perhaps, to describe the central

43. Cf. PI §593.

44. Lamarque and Olsen, *Truth, Fiction, and Literature: A Philosophical Perspective*, ch. 13.

45. Kivy's reply is that "Some fictional works contain or imply general thematic statements [i.e., theses] about the world that the reader, as part of an appreciation of the work, has to assess as true or false" (Kivy, *Philosophies of Arts: An Essay in Differences*, 122).

46. Lamarque, "Literature," 571.

aims of literary criticism. We wouldn't, however, do so well using it to characterize much everyday literary appreciation. First, using it, literature looks essentially dour, and, were this so, then, insofar as it entertains us, we're not appreciating it, and, insofar as we recommend it as funny, we're being frivolous. Second, using it, literature looks essentially like a study of language, and, were this so, then, when we read mainly for the story, we hardly appreciate it, or, if we recommend a novel as a page-turner, we badly undersell it. My point isn't that, by the light of everyday appreciation, the analysis too sharply distinguishes literary from other writings. Rather, it's that such appreciation extends from the heights of such things as beautiful language and a serious moral purpose down into the lower realm of entertainment and its sundry kin.

But are there works of literary fiction popular enough to be so appreciated? Yes, for example, well-known novels in and around the Western Canon—such as *War and Peace*, *1984*, and *Infinite Jest*—and these are those I'll focus on here, for two reasons. One is that they're usually the ones cited in the philosophical literature to which I'm responding. In line with this, the other is that they are, because, e.g., "imbued with moral seriousness," more apt to be, or to be regarded as, thought experiments than other popular writings, such as the expressly non-literary action and adventure novels of a Clive Cussler or a "Fifty Shades" romance by E.L. James. Now, to be explicit about what, for the most part, I'm leaving aside, it includes such works as Shakespeare's plays, Homer's epics, Aesop's fables, and Chekov's short stories. Also, I do so even though these works resemble thought experiments in ways that novels do not and, so, themselves deserve philosophical treatment; think, for instance, of the Bard's sonnets, especially their thought-experiment-like concision, imagery and concluding couplet.

Now, the expression, "everyday appreciation of literature," looks like an oxymoron. After all, it's no everyday matter but one of serious study to read Joyce's *Ulysses* and "appreciate it," insofar as this means that we "apprehend or understand clearly or correctly" or "recognize or grasp the significance or subtleties of" the work.[47] Is the expression then even coherent? Insofar as "appreciation" means, in significant respects, to so apprehend or recognize the work, yes, it is. And this is how I mean it. Put another way, I treat mastery of a work of literary fiction as an ideal that everyday readers approximate to some limited extent.

Finally, to elaborate a little on this expression, when I use it, I have in mind what the reading public gets out of the works doing what they generally do with them, especially interpreting and recommending them or else discussing them in book clubs or perusing reviews or a biography, and so on. For example, in addition to reading David Foster Wallace's *Infinite Jest*, you might look up online annotations to interpret difficult passages, e.g., to get at allusions, broad themes and subtleties of the plot; read reviews to gain insights for writing about or discussing the book's virtues or faults or for showing an understanding of its significance when recommending it; or place the book in the context of the author's life by perusing his more accessible long-form journalism, listening to a biography, and watching a Hollywood biopic.

***

Here is the third and final worry. By focusing on everyday appreciation instead of mastery, don't I miss out on useful objections? To a limited extent, yes. To explain, often, for writers and readers of literary fiction, tight-knit integrity of the work is an important ideal. For example, we'd like to say that, to really understand a story in such a work, one must take the whole work into account, and that, consequently, if we cut a story out of the work and discard the rest, we do not really understand it. Alternately, to defend a novel's length, we say every bit of it is necessary. For example, in Dave Eggers' Forward, "The Book *Infinite Jest* is 1,079 pages long and there is not one lazy sentence."[48]

47. *Appreciate v. 3a*.
48. Wallace, *Infinite Jest*, xii.

By the light of such an ideal, one may argue that we fail to really understand stories in works of literary fiction if we either take them to be or regard them as thought experiments. For example, if, from Styron's *Sophie's Choice*, we cut out the novel's namesake scene,[49] and treat it as a thought experiment in an ethics class, one may argue that, in so doing, we do not really understand the story, since it is what it is in virtue, among other things, of its connections to the rest of the work.

To be sure, I don't take this to be some knock down argument. After all, couldn't we, when merely speaking of a story in a work, be dealing with the whole thing? Perhaps. For example, when we take the choice scene from Styron's novel to be a thought experiment, one might argue, we may normally albeit implicitly take into account the rest of the work, as Martha Nussbaum does explicitly quoting a passage from James' *Golden Bowl*.[50] Still, the objection has much potential force.

I cannot make such an objection to approaches like Davies' and Elgin's, and this is a cost of focusing on everyday appreciation instead of mastery. That is, I'm already dealing with appreciation that doesn't "really understand" the stories, and so I cannot object that their accounts fail because, on them, readers fail to so understand. To this limited extent, because of my focus, I miss out on what would be useful objections.

3.2.1 Two Difficulties for Davies' Route

In this section, after summarizing Davies' route, I explain the above two difficulties specific to it. The second difficulty touches on those three disanalogies, which also underlie my criticism of Elgin's account. This criticism I explain afterward as well as in the following two sections, where I further explain the disanalogies.

Davies' Route

In his argument for an identity claim, Davies makes four moves. First, he offers two individually necessary and jointly sufficient conditions on something being a narrative fiction. One, borrowed from Gregory Currie's make-believe account of fiction,[51] is that its author intends that "we make-believe, rather than believe" the story narrated.[52] (On what "make-believe" means, see the first criticism below.) The other, taken from his own work,[53] is that its author's primary aim not just be to relate events that occurred in the order they occurred.[54] On this account, for example, *War and Peace* counts as narrative fiction for two reasons: Tolstoy intended that his readers make-believe the story he tells, which is of events beginning before Pierre's return to Russia and ending around his marriage to Natasha; and, he meant not only to tell his readers a true story. For instance, he wanted, in writing about certain French invasions, not, or not only, to relate a sequence of historical events but, in light of his epilogue, to show figures like Napoleon not as chess players moving pieces but as mere pieces being moved.[55]

Davies' second move is to argue that, since thought experiments satisfy these two conditions, they're narrative fictions. To begin, he states that a thought experiment narrative has three parts.[56] First, we're presented with a hypothetical situation, which contains a process or event. For instance, in Jackson's well-known Mary thought experiment,[57] there's a hypothetical situation in which Mary-the-colour-scientist

49. Styron, *Sophie's Choice*, 526–530.
50. Nussbaum, "Finely Aware," 149.
51. Currie, *The Nature of Fiction*.
52. Davies, "Thought Experiments and Fictional Narratives," 31.
53. Davies, "Fiction."
54. Davies, "Thought Experiments and Fictional Narratives," 31.
55. Tolstoy, *War and Peace*, 1215–1225, 1077–1078, 1314.
56. Davies, "Thought Experiments and Fictional Narratives," 32. This tripartite structure has many precedents, e.g., in the introduction to Gendler, *Thought Experiment: On the Powers and Limits of Imaginary Cases*.
57. Jackson, "Epiphenomenal Qualia."

perceives the colour red for the first time. Second, there's an outcome of the hypothetical situation's process or event, such as Mary, despite all she knows, nevertheless learning something. Third, there's the outcome bearing on a more general issue, for instance, Mary's learning purportedly showing there to be more than material objects and relations between them. This tripartite division made, he argues that thought experiments satisfy the first condition. His premise, which he finds plausible, is that a thought experiment's author expects us to make-believe—rather than believe—the hypothetical process or event.[58] For instance, if so, Jackson intended that we make-believe—as opposed to believe—that Mary saw the colour red for the first time. Thereby, Davies precludes the possibility that we entertain an event which we believe to have happened, which I'll argue below raises one difficulty. Finally, he argues that thought experiments satisfy the second condition. That is, since their authors do not think such hypothetical situations and outcomes actually occurred, they do not aim primarily to relate how these events occurred in the order they occurred.[59] For example, since Jackson didn't think the hypothetical Mary events and their outcome occurred, he didn't aim to so relate them, much less aim primarily to do so.

Third, Davies, replying to an objection, denies that a certain difference, one like those we saw Mach point out above, divides the authors of narrative fictions from those of scientific thought experiments. The purported difference, in short, is that, whereas authors of fiction don't intend what occurs in their stories to occur as it really does, those of thought experiments do.[60] To illustrate narrative fictions that don't have this aim, he gives as examples two genres and two works of literary fiction: "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."[61]

Fourth and finally, given this no-difference in authorial intention and that his answer to his own question—"To what extent are TE's (like) fictional narratives?"[62]—is that they are entirely so, he concludes: "Perhaps, then, we should simply allow that some works of fiction are properly viewed as much more fully elaborated TE's."[63] This conclusion is the identity claim he makes. Even if true, one cannot infer from it that any works of literary fiction—normally so-called—are thought experiments, and this gives rise to the second difficulty, discussed below.

First Difficulty for Davies' Argument

Davies' argument depends on this conditional: A thought experiment is a fictional narrative only if its author expects or intends us to make-believe its narrated story. To see a consequence, recall the contrast between "make-believe" and "believe." This contrast works such that if you make-believe that, say, you're an astronaut, then it's not the case that you believe it. More to the point, it works such that, if you expect or intend us to make-believe we're astronauts, then, if we *believe* that we're astronauts, we're not doing what's permissible, i.e., what you expect or intend us to do. The consequence is that a thought experiment is not a fictional narrative, if we're permitted to believe its narrated story.

58. Davies, "Thought Experiments and Fictional Narratives," 32.
59. Davies, 32.
60. The original objection, Nancy Nersessian's, is, more specifically, against scientific thought experiments being fictions (Nersessian, "In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling," 295–7). Also, Davies considers two other accounts of the intention. They are that readers *believe* the objects so behave and that *readers believe in order to make-believe* that they so behave (Davies, "Thought Experiments and Fictional Narratives," 33). His response in light of each alternative is essentially the same plausible denial. In short, it is to deny that all fictions lack the intention.
61. Davies, 33.
62. Davies, 31.
63. Davies, 33.

The difficulty is that, at least sometimes, we are permitted to believe it. To see that this is so, recall Galileo's falling bodies thought experiment. Its "story" comprises, minimally, a large freely falling stone attached to a smaller one. It may have happened. Suppose you believe it did. Even so, believing the story, you can, if differently, nevertheless carry out the thought experiment. For example, you may suppose the stones are falling, then recall that actual ones did fall or else not recall the belief, and carry on, i.e., ask yourself whether they fall faster than the large one alone, and then reason that for Aristotle they fall both faster and slower, and so on. Nothing hangs, in the thought experiment, on whether you believe the case actually occurred or not. So Galileo, the author, shouldn't be taken to expect or intend that we make-believe the stones fall; rather, we should take him to permit the belief that they do—e.g., take him to presume that our belief (or disbelief or simple non-belief) makes no difference whatever.

Now, the opposite, that we're not so permitted, may have seemed true because of a confusion of scope. That is, we may have confused there being no need to believe what we suppose, which is true, with there being a need to not-believe what we suppose, which is false.

But is Davies *committed* to the view that make-believe precludes belief? Perhaps, since, with various qualifications, he accepts Gregory Currie's well-known make-believe theory of fiction.[64] After all, as Currie uses it in the theory, "'Make-believe' is not a term of art";[65] and, the term's ordinary use implicates non-belief.[66] That is, often as a matter of course, to say "S make-believes p" implicates that "S does not believe p." To illustrate, consider this directive: "make-believe that your pen is a sword, and believe it too." Compare this to G.E. Moore's "it's raining outside, but I don't believe it." In both, a "pragmatic contradiction" arises. That is, in the latter, I got across that I believe it's raining, but then I say that I do not believe that it's raining. Likewise, in the former, I got across that I want you not to believe that your pen is a sword, but I also tell you to believe it is a sword. Now, to avoid this difficulty, couldn't Davies simply use another make-believe theory, in particular, Kendall Walton's? No, not simply. After all, Walton too seems committed to an ordinary notion of make-believe. We see it, for example, in his claim that "The activities in which representational works of art are embedded and which give them their point are best seen as continuous with children's games of make-believe."[67]

Second Difficulty for Davies' Argument

As we saw, the fourth and final move Davies makes is an inference. Schematically, it's that, since thought experiments are narrative fictions, some narrative fictions are thought experiments. I've just tried to undercut the premise, but, for the sake of argument, assume it is true. The difficulty I want now to point out is that, although this move makes an end run around the difficulties of relying on a definition of thought experiments, or the like, it doesn't get him what he wants, namely that certain stories in works normally recognized to be literary fiction are thought experiments. This is so whether we construe it deductively or inductively.

If, less charitably, we construe the argument as deductive, it has the form *All Fs are Gs* therefore *Some Gs are Fs*. *All Fs are Gs* obviously doesn't entail *some Gs are Fs*, on account of the possibility that there are no Fs; but, in this case, there are Fs, since the Fs are thought experiments. Therein our problem does not lie, but it is nearby. To see where, distinguish between thought experiments that are narrative fictions and the narrative fictions we already call "stories in works of literary fiction," such as those in Orwell's *1984*. The conclusion,

64. Cf. a central thesis in Currie's *The Nature of Fiction*, "that we can define fiction itself in terms of the author's intention concerning our make-believe" (Currie, *The Nature of Fiction*, 18).

65. Currie, 19.

66. Cf. Christopher New's argument (New, "Walton on Imagination, Belief and Fiction," 160) & John Gibson's approval of it (Gibson, *Fiction and the Weave of Life*, 164-170).

67. Walton, *Mimesis as Make-Believe: On the Foundations of the Representational Arts*, 11.

schematically, that some narrative fictions are thought experiments, should imply that some narrative fictions qua stories in works of literary fiction are thought experiments, but it doesn't. It doesn't because it would be true even if none of the stories were. It would be true even then because the thought experiments we're calling "narrative fictions" are thought experiments.

Let us now, more charitably, construe the inference as inductive. One plausible construal, in light of the conclusion's qualifications, is that the argument is supposed to be analogical and burden shifting. Such an argument might run as follows. Thought experiments have all the properties needed to be narrative fictions. So works of fiction, such as dystopian novels like *1984* and *Brave New World*, have many properties in common with thought experiments. Indeed, they have so many that the burden of proof seems to lie with anyone who doesn't hold that such works of fiction are thought experiments. Thus, tentatively, we should take some works of narrative fiction to be thought experiments. My objection consists in pointing out three dissimilarities, i.e., disanalogies, which shift the burden back. They are, in short, that the stories in such works, unlike thought experiments, may be complex, can be interpreted freely, and are for imaginative flights of fancy. I'll outline them here and, in the following two sections, elaborate.

When we begin to contrast the stories with thought experiments, a difference in complexity comes to mind. As we saw, this difference moves Sorensen to claim that one isn't the other. That is, he argues, in short, that the stories, because complex, don't aim primarily to answer theoretical questions, and so, by his definition, aren't thought experiments.[68] Against such a line, as we'll see in §3.4.2, philosophers such as Davies, Elgin, and Noël Carroll hold that this difference hardly precludes one from being the other; rather, complexity improves thought experiments. But, crucially, we call the stories complex, not only because they're detailed, but because they are often hard to take in, to recall, and to describe—unlike thought experiments. Indeed, as a matter of course, we often recognize thought experiments, unlike the stories, in virtue of their being surveyable. That is the first difference.

That the stories lack conclusions also readily comes to mind when contrasting them with thought experiments. That is, when we appreciate the stories, we are, as a matter of course, permitted much "free-floating contemplation";[69] conversely, when we understand a named thought experiment, we cannot, as a matter of course, take it to have any conclusion other than that explicitly given. This point faces a powerful objection in Michael Bishop's influential criticism of accounts of thought experiments like John Norton's, as I'll discuss and reply to in §3.4.1. In sum, the stories permit a freedom of interpretation which thought experiments do not. That is the second difference.

Finally, that the stories are flights of fancy also comes readily to mind when contrasting them with thought experiments. Mach, as we saw in the introduction, nevertheless allows novels to be thought experiments, although not those of serious enquirers—and this exception comes at the cost of denying that we learn from the novels. Davies, by contrast, proposes counterexamples, i.e., non-fanciful works of dystopian and utopian narrative fiction such as *1984* and *Brave New World*. Perhaps Zola would have added to Davies' list "experimental novels" like Balzac's *Cousine Bette*. Nevertheless, if, as I'll argue in §3.3, we want to understand learning from stories in everyday literary appreciation, certain differences remain—ones between some interpreting and recommending uses of story imaginings and certain obervation-like thought experiment ones. This is the third difference.

I'm arguing here that these three disanalogies shift the burden back and so present a difficulty for Davies' argument, if we take it to be analogical and burden shifting. I won't, by the way, construe his argument as

68. Sorensen, *Thought Experiments*, 223.

69. Cf. "[Atget's crime-scene-like photographs of Paris] demand a specific kind of approach; free-floating contemplation is not appropriate to them" (Benjamin, *The Work of Art in the Age of Mechanical Reproduction*, §VI).

inductive in any other way. That said, if done, I suspect the three disanalogies would, as differences, nevertheless raise problems. For example, if the argument were construed as abductive, they may well challenge the identity claim as a best explanation, or, if as a generalization from dystopian and utopian narrative fiction taken to be thought experiment, they may well challenge either the base case or the inductive step.

In light of this difficulty, as well as the preceding one, Davies doesn't justify his identity claim. If we follow him, then, "we risk losing our grip on the stories," as explained above. This is one half of my central contention.

3.2.2 Difficulties for Elgin's Route

The other half, again, is that, in this way, we also risk losing our grip if, following Elgin, we merely regard certain stories in works of literary fiction as thought experiments. This resembles a criticism Davies makes of Elgin's account, which I'll consider in §3.4.2. For now, I'll explain my criticism, and, in what follows, I'll develop it.

Bypassing identity claims, Elgin cannot simply apply an account to thought experiments and thereby apply an account of thought experiments to certain works of fiction. She doesn't try to. Rather, she applies one and the same account to experiments, to thought experiments, and to fictions. As she puts it, summing up her view, "Whether or not we call works of fiction thought experiments, I have urged that fictions, thought experiments, and standard experiments function in much the same way"; and, she goes on to specify the sameness:

> By distancing themselves from the facts, by resorting to artifices, by bracketing a variety of things known to be true, all three exemplify features they share with the facts. Since these features may be difficult or impossible to discern in our everyday encounters with things, fictions, thought experiments and standard experiments advance our understanding of the worlds and of ourselves.[70]

Now, to urge us that each functions in this same way, Elgin argues in different ways. Specifically, she argues *first* that experiments do so, *then* that thought experiments do, and *finally* that fictions do. The ordering, as I read the argument, isn't trivial. Rather, we're to apply the account to fictions in light of applying it to thought experiments, which we do in light of applying it to experiments. This first in-light-of application, put another way, is that, to apply the account to fictions, we are, among other things, to regard them as thought experiments. If this seems to be a leap on my part, notice that Elgin argues that we can so construe them, as I will touch on in §3.4.2. Now, in so regarding them and applying the account, we risk losing our grip on what we want to explain. We see this once we take the above three differences into consideration. That is, as I'll argue below, explaining as Elgin does, we inadvertently replace the fictions' complexity with surveyability, constrict their interpretive freedom, and ban certain flights of fancy imagining them; and, thereby, we risk explaining learning from a lookalike and not from stories in works of literary fiction ordinarily appreciated.

3.3 Imagination Differences

I will now begin to elaborate the three differences, sketched above, on which my argument primarily depends. In this section, I explain how the imaginings we have appreciating stories in works of literary fiction are, as a matter of course, for flights of fancy, and so how they differ from thought experiments' observation-like ones. In particular, unlike these observation-like ones, literature's flights have as proper ends in themselves such things as literary effects, entertainment, and story recollection. In light of this section, in the next, I'll give a comparatively brief explanation of the other two differences. These two concern, instead of imaginings,

70. Elgin, "Fiction as Thought Experiment," 240.

outcomes and overall structure. Specifically, they concern, instead of flights of fancy, interpretive freedom and complexity.

My explanation here has two main parts. First, I set up for an account of how we ordinarily take imaginings when appreciating works of literary fiction. To do so, I distinguish imaginings that are experiences of the work from those we merely have while and because of reading it. That done, I can ask how we take the imaginings we have to be experiences of the work. Then, I outline a partial answer and fill it in. To fill it in, I sketch how, interpreting and recommending the work, we so take the imaginings. In these sketches, we see those just-mentioned proper ends of using imaginings, i.e., ones for literary effect, entertainment and story recollection. With this partial answer in hand, in the second main part, I explain the differences. That is, I describe how, in certain respects, the ways we normally take our imaginings in such interpreting and recommending differ from the ways we would take them were we either, like Elgin, to regard the work's stories as experiments in thought or, like Davies, to identify the ones with the others. Again, along the way, I develop my criticism of Elgin's route.

3.3.1 How We Take Imaginings To Be Experiences of Literary Fiction

Read through the following excerpt from Joyce's *Ulysses*:

> He came over to the gunrest and, thrusting a hand into Stephen's upper pocket, said:
>
> —Lend us a loan of your noserag to wipe my razor.
>
> Stephen suffered him to pull out and hold up on show by its corner a dirty crumpled handkerchief. Buck Mulligan wiped the razorblade neatly. Then, gazing over the handkerchief, he said:
>
> —The bard's noserag! A new art colour for our Irish poets: snotgreen. You can almost taste it, can't you?
>
> He mounted to the parapet again and gazed out over Dublin bay, his fair oakpale hair stirring slightly.
>
> —God, he said quietly. Isn't the sea what Algy calls it: a great sweet mother? The snotgreen sea. The scrotumtightening sea. *Epi oinopa ponton.* Ah, Dedalus, the Greeks. I must teach you. You must read them in the original. *Thalatta! Thalatta!* She is our great sweet mother. Come and look.
>
> Stephen stood up and went over to the parapet. Leaning on it he looked down at the water and on the mailboat clearing the harbour mouth of Kingstown.[71]

Reading through this passage, what did you experience?

Wasn't it more than words on the page, ones such as "razor," "wipe," "noserag," "gazed," and "snotgreen sea"? Among other things, wasn't there also what you felt and imagined? Two examples. I felt queasy at "… snotgreen. You can almost taste it, can't you?" Now, to feel while reading, that's unremarkable; we hardly even notice experiences so very familiar. This queasy feeling, however, because of its strength, did surprise me. Moreover, were I asked to explain my surprise, immediately to mind would come *disgust*, that of tasting, or the taste of, snot—not to mention *green* snot—and I'd answer, grimacing, "It's as if I tasted it." That is, while reading, beyond the words, I had a "gustatory imagining," one striking because disgusting. That's the first example. Here is the second, a mercifully less evocative one. If asked what I remember reading the words, "The snotgreen sea," immediately to mind comes a visual imagining, a greenish watery one, and one I recall as the snotgreen water of Dublin Bay. Now, like feeling while reading, to visualize something while doing so raises no eyebrow, but the novel and unpalatable colour comparison raises both brows—all the more so juxtaposed as the comparison is with a romantic metaphor, i.e., that the sea is "a great sweet mother," attributed to the poet Algernon Charles Swinburne. That is, while reading, beyond the words, I had a "visual imagining," one

71. Joyce, *Ulysses*, 5.

surprising because new and slightly gross. Again, beyond the words, didn't you experience feelings and imaginings, such as these unsavoury visual and gustatory ones?

I assume you did, but, even if not, I can still make my point. It is that, imaginings like these, they are experiences *of the literary work.* And we take them to be so. I want to ask how? How do we take evocative imaginings like those, had while and because of the work, to be of it?

But can we ask it? You might think not. That is, you might think there's no logical space for such a question, since to have such imaginings while or because of reading the work *just is* to take them to be experiences of it.

But to so have them isn't to so take them. First, to be an experience of the work isn't simply to be had *while* reading it. To hungrily imagine your ice-cream-stocked freezer while reading isn't, for example, to experience the work, usually at least. By analogy, to notice background noises while reading, likewise, isn't to experience it, except in special circumstances, e.g., when we're easily distracted because trudging rubber-booted through a swamp of turgid prose, thick with obscure *Agapē Agape*-like allusions and constantly OED-requiringly recondite diction. Second, to be such an experience isn't simply to be had *because* of reading the work. For instance, to imagine a snotty queen because of both reading "snotgreen" and noticing a similarity between this word and the words "snotty queen" isn't to experience it, usually at least—much like reading-induced eye strain isn't, unless it's due, say, to our being so gloriously absorbed in the work that we read feverishly all night long. Finally, third, to be an experience of the work isn't simply to be had *both* while and because of reading the work, since to imagine the snotty queen mid-sentence isn't to do so. In sum, not all imaginings had while or because of reading a work of literary fiction are ones we take to be experiences of it, and, consequently, our so taking them isn't simply our so having them.

Whence the inclination to think there's no logical space for the question, even though there is? It may arise from the effort required to think of cases in which we do not so take such imaginings. To see this, recall the line: "He mounted to the parapet again and gazed out over Dublin Bay, his fair oakpale hair stirring slightly." I find it *easy* to recall feelings and imaginings had while and because of reading this—e.g., the uplifting feeling had imagining pale oak-brown hair stirring slightly or that sense of vastness felt imagining Dublin Bay gazed at from upon high—and I take such imaginings and feelings, had while and because of reading *Ulysses*, to be of it. By contrast, I find it *hard*, but not impossible, to recall experiences had while and because of reading the work which aren't of it. After all, first, I'm usually unable to recall ordinary ones, e.g., of computers humming, of my left to right eye movement, and of my jaw's tightness or slackness. Second, I have to put in some effort merely to invent such an experience, especially one that isn't contrived. An example, a contrived one: While reading the line above and because of it, I might feel hungry. Perhaps the word "gazed" sounds like "grazed" and I imagine a cow grazing. This experience of hunger, for all that's been said, clearly isn't of the work—and to think up this possible not-of-the-work experience takes some effort. That is, these imaginings, which we have while and because of reading the work but which we do not take to be of it, take some effort even to think of; whereas we easily recall those so had but which we do take to be of it. If we overlook such effortful imaginings, we may be inclined to think that to have imaginings while or because of reading the work *just is* to take them to be experiences of it. Thereby, the effort may explain the inclination—after reading, "How do we take such imaginings to be of the work?"—to think this question vacuous.

But can't we ever take imaginings to be of a work in virtue of having them while or because of reading it? I don't wish to deny that we sometimes do. Again, I aim only to show my question askable, and, to this end, I only deny that to so take them just is to so have them.

To sum up, we take evocative imaginings, such as the snot and oakpale hair ones, had while and because of

reading the work, to be of it. But, as we saw, so having them isn't what it is to so take them. So, we can ask: How do we so take them? That is, in virtue of what if anything do we take imaginings like those, ones had while or because of reading a work of literary fiction, to be experiences of it?

Partial Answer to: How Do We Take Imaginings To Be Experiences of Literary Fiction?

A partial answer will suffice for my purposes. That is, to answer, I will describe only some of the ways in which we take imaginings to be of a work of literary fiction when appreciating it. In particular, I'll describe those ways that, as I'll explain below, differ from how we take them when carrying out a thought experiment.

This answer has the form *we often do so as in these examples*. Filling the form in a little, I'll answer that we often take imaginings, had while and because of reading works of literary fiction, to be experiences of them in virtue of what we do with them, as when we interpret and recommend. To further fill it in, I'll sketch these interpreting and recommending uses.

To set up for this, I'll first explain three things: how such a description can answer; what it is to "use" imaginings; and why in certain ways all this shouldn't worry us. After the answer-filling sketch, I'll be in position to explain the disanalogy—i.e., describe how these uses of imaginings differ from what they would be were their stories regarded as or taken to be thought experiments.

How My Description Could Answer

Again, in this answer, the examples of interpreting and recommending do not stand for every possible literary use of our imaginings. In particular, they're used only to describe and not, in line with my first chapter's approach, to advance a theory of such uses. Such a theory, for example, might include a hypothesis, one which posits a commonality, and one meant, if confirmed, to explain how imaginings count as experiences of literary fiction. Without doing so, however, how could the description I give at all answer the question?

Well, it does so much like, by describing basic positions and moves in chess, we teach or remind someone how to play. That is, the imaginings we recall having while interpreting or recommending a work of literary fiction are like pieces in chess, and, as we can teach or remind someone how to play chess by describing possible moves and positions, we can, to explain how we take imaginings to be experiences of literary fiction, describe certain roles these imaginings have in our interpreting and recommending.

But can we give such a chess-like description? After all, no chess-like system of rules governs what, in interpreting or recommending a work, we are to do with what we imagine reading it. Admittedly, literary "rules" aren't systematic like those in chess. I can, for example, be recommending *Ulysses* even if I appeal to its gross greenish imagining to do so, but I cannot be playing chess if I always move my knights as queens do. I may, instead, of course, be playfully testing a child's understanding of the game. This difference, however, isn't sharp. After all, first, we cannot rightly recommend based on just any imagining, and, second, we do not always cease to be playing chess after breaking a rule of the game. For example, a recommendation based on imaginings not had while or because of reading the book will usually miss the boat, and, usually, we rightly albeit unreflectively call games "chess," even if the players inadvertently omit or don't know special rules such as en passant capture, e.g., if they're kids having fun outside a formal competition.

How Using Imaginings Outstrips Both Merely Having Them & Preparing to Use Them

The partial answer accounts for certain imaginings being of certain works by appeal to their use. But what is it to use them? To explain, I'll ask and answer two questions.

First: What is it to use them as opposed merely to be having them? To press the question, by analogy, you need not use the hammer you have. You simply hold it without hammering. Then, to use it as a hammer, you hammer. Similarly, you can have an imagining of a hammer. You imagine a hammer. But to use the imagined hammer can be nothing other than to have another imagining—specifically, to imagine hammering. You can't, after all, use an imaginary hammer as you could a real one. This, of course, isn't to deny that you can imagine having a hammer and then imagine using that hammer, e.g., have an imagining of a hammer in hand which you merely hold and then use to nail a nail. Here is the point. In this case, the distinction between having and using is without a difference. Why then isn't the having/using distinction I draw on also one without a difference?

Suppose you say something like, "the snotgreen sea I imagined visually is an experience of the novel *Ulysses*," and you can recall having the imagining while or because of reading the work. So far, you haven't yet done anything with the imagining. You've made no move. You may be preparing to do so, like naming something to talk about it. For example, having said it, you may dissuade a potential reader, saying, "Don't read it. That green-snot imagery befouls literature!" Dissuading in this way, making such a move, you use the imagining, referring to its purported foulness to make your point. To be sure, this isn't like using an imagined hammer insofar as such "using" isn't anything if not imagining hammering. After all, you're dissuading, not imagining dissuading. So the having/using distinction I want to draw does indeed have a difference, which prevents my answers from falling into nonsense.

Now, second, let us ask: What is it to use them as opposed to be preparing to use them? After all, isn't setting up such a connection itself not having but using the imagining, again washing out the distinction's difference? No. This connecting isn't "using" an imagining as we applied the expression, i.e., isn't doing with what is used what it's for. By analogy, we take the hammer out, get nails ready, and choose where to hammer them, all of which isn't yet hammering but preparation for it, not *using* a hammer, even though you can call all of this "using a hammer." That is, calling it all "using a hammer" would be correct insofar as you did something with the hammer but, for the most part, would also mislead, since we generally take the expression to mean hammering and not preparation for it. Finally, if you like, the difference may be characterized as one between doing something with the imagining and doing with it what it's for, but then, again, it doesn't wash out.

In sum, there is having an imagining while or because of reading the work; preparing to use it; and, using it. It is certain uses of such an imagining that, I'm claiming, often make it an experience of the work. By analogy, in virtue of certain L-shaped movements, a bit of wood is a knight in chess.

Assuaging Three Worries

Again, I'm not saying that a use necessarily makes it so. By analogy, moving a lowly pawn L-wise doesn't raise it in the world to knightly heights. After all, using an imagining had improperly while or because of reading a work doesn't raise it up to an experience of that work. Suppose I imagined a hot green sea, misreading "snot green," and used my imagining to "evaluate the work." You would be right to deny that I so used it and, consequently, that my imagining even counts as an experience of the work.

This helps us deal with our first worry: Aren't experiences of a work, unlike those of chess, entirely subjective? That is, isn't any experience I think I had of the work *my experience of it* or properly had—whatever you might think? They aren't, at least insofar as I can, with your help, discover my own misreading. For example, if I find that I misread "snot green"—and I may come to think so because you pointed it out to me—I ought to change what imaginings I take to be my experience of the work.

That was one worry. Another is that I'm closing off a certain reverse avenue of explanation, i.e., preventing

us from explaining interpretation or recommendation *in terms of* our experiences had while or because of reading the work. But, again, I'm not advancing a theory, and, in particular, I'm not fixing an explanatory direction. For example, I need not deny that you can explain how you took Raskolnikov's redemption in *Crime and Punishment* by appeal to what you visually imagined reading the novel's end. After all, I allow that these imaginings may be of the work other than by their interpretive use.

The third and final worry is that, contra my account, some unused imaginings and some merely illustrative ones are of works. To assuage the worry, take each sort of imaginings in turn. First, consider unused ones. If an author uses concrete language merely to evoke visual imaginings in readers, for instance, we may take those imaginings to be of the work even if we don't use them. But this doesn't conflict with my account. I don't claim that a reader's uses alone make imaginings of the work, and I don't deny that an author's intentions might also do so. Second, consider merely illustrative imaginings, those like the pictures in a children's book. When we rightly read a work's words and imagine something in line with them, but do not use the imaginings to do anything like interpret or recall the work, the imaginings may well thereby count as experiences of the work. Suppose, for example, I'm reading *JR* by William Gaddis and I recognize a line from Dickens' *A Christmas Carol* and suspect an allusion. Perhaps that prompts me to compare characters in one work to the other. Doing so, I might match the roles of the characters Eigen, Bast, Gibbs and Schramm to Scrooge and the ghosts of Christmas past, present and future. That done, I might interpret the Dickens allusion as follows. It means that Eigen has been warned to turn from money back to art, as Scrooge back to generosity, and, also, that the novel's theme isn't simply that money corrupts art. Now, consider the imaginings I might have while reading and interpreting, e.g., of a faucet running, of Ostrich eggs, and of copulation seen through an apartment window. They illustrate what I read but fly free of my interpretive activity. Yet these imaginings may be experiences of the work. But, again, this doesn't conflict with my account, since I don't claim that only interpreting and recommending uses of imaginings make them of the work, and I don't deny that merely being illustrative could also make them so count.

Uses of Imaginings in Recommending & Interpreting

So far I've outlined my partial answer. Again, it is that we often take imaginings, had while and because of reading works of literary fiction, to be experiences of them in virtue of how we use them, e.g., when interpreting or recommending. Now, for all I've said, how we take imaginings in these works hardly differs from how we take those in thought experiments. After all, we often have imaginings while or because of carrying out a thought experiment, and these imaginings need not be experiences of it. For example, if, while or because of reading about the two balls in Black's thought experiment, as discussed in §2.3.1, we imagine not spheres in empty space but various glamorous Russian balls of the 1800s, complete with gowns and the Mazurka, we have imaginings that aren't of that thought experiment. We can then ask how we take imaginings had while or because of a *thought experiment* to be experiences of it—and we might give, as a partial answer, that we do so in virtue of how we use them. How we do so, however, as I'll now explain, differs from how we use them interpreting and recommending works of literary fiction.

To do so, I will, here, fill in the partial answer I've outlined by sketching certain recommending and then interpreting uses. Crucially, these uses are for literary effect, for entertainment, and for story recollection—and they aren't merely means for some other end. Afterward, I will contrast these uses with a central one in thought experiments.

To help you see where I'm going, let me add—in light of last chapter's survey of affinities, in §2.2.3—that this central use of imaginings in thought experiments has experimental observations as an ideal. For example,

we carry out Black's thought experiment ourselves and use our imaginings of the spheres as objects about which to reason—as we do observations in experiments—aiming either to confirm or disconfirm a certain proposition. Alternately, and more realistically, we read his thought experiment, thereby acquainting ourselves with those sphere imaginings, then follow his reasonings about them to the conclusion he draws, as we might do reading an experiment report, and, then, perhaps, we use the imaginings ourselves, as if they were our own observations, to evaluate his reasoning and the conclusion he prescribes—as if repeating an experiment to check it. It is this observation-like use, as we'll see, that differs from the interpreting and recommending ones I'll now sketch.

A Sketch of Recommending Uses

To recommend a work of literary fiction we often appeal, as a matter of course, to what's good for both the head and the heart or, as Davenport might put it, to both cognitive and emotional development. Imaginings often underlie these appeals. Examples abound.[72] Consider one. It's a recommendation arising from an interpretation of David Foster Wallace's *Infinite Jest*.

We might interpret the book, first, in light of his widely read Kenyon College Address, to concern, among other things, both directing our attention and empathy.[73] Second, we may interpret it, in light of what he says in interviews, along two lines: along one, that it's intended to be "fun enough so that somebody would be almost sort of seduced into doing the work;"[74] along the other, that it's intended to help us, as it were, leap the wall and know other people.[75] In sum, on this interpretation, the novel, among other things, by being fun, offers its readers a means both (i) to work on learning to know and to empathize with others and, thereby, (ii) to better cope with loneliness. More specifically, the entertainment in certain characters' stories—which imaginings about these stories supports—helps us to pay attention to them. Doing so, we notice, again and again, across varying cases, how they see yet other characters' stories as like their own. Thereby, we also notice

72. We recommend: William Gaddis' *A Frolic of His Own* and *JR* as sharp and hilarious satires, of law and business respectively—one imagining they share being that of *Cyclone Seven*, a piece of public art that entraps, in one novel, a child (Gaddis, *JR*, 671–672), and, in the other, a puppy (Gaddis, *A Frolic of his Own*, 29 ff.), and that calls out for interpretation as a symbol; Dostoevsky's *The Brothers Karamazov* for its psychological truths and carnivalesque feel, the supportive imaginings including that of the devil Ivan hallucinates, vividly described as a gentlemanly moocher who'd like best to be incarnated as a merchant's wife and who discusses other eschatological topics with him (Dostoevsky, *The Brothers Karamazov: A Novel in Four Parts with Epilogue*, 628–644); or Tolstoy's *Anna Karenina* as a mixture of tragedy and philosophy, the imaginings including those had reading the novel's climactic event, horrified, as Anna regrets throwing herself onto the tracks, and then thoughtful, recalling this event's foreshadowing and, perhaps, seeing in it an idea about the sealed and unjust fate of "fallen woman" (Tolstoy, *Anna Karenina*, 768); or Proust's *Remembrance of Things Past* as an intellectual tour of largely bourgeois and aristocratic French life in the decades leading up to the First World War, the imaginings including those that introduce themes early in the series, such as that of decline, which is introduced by the striking image of the narrator's grandmother, vividly and humorously or else poignantly described, walking alone outside *because of* the rainy weather, while the rest of the family, having no such hearty ardour, remains inside teasing her (Tome 1, Première Partie, Combray I, Proust, *À la Recherche du Temps Perdu*); or Austen's *Pride and Prejudice* as, by her own lights, especially witty (Elborough and Gordon, *Being a Writer: Advice, Musings, Essays and Experiences from the World's Greatest Authors*, 126) but also, presumably, of good sense, and to mind may come imaginings had reading of Caroline Bingley inviting Elizabeth Bennett to "take a turn about the room," which sets up a delightful but also thought provoking scene, one about moral emotion, namely, that in which Darcy says, "it has been the study of my life to avoid those weaknesses which often expose a strong understanding to ridicule," and in which Elizabeth acutely replies, "Such as vanity and pride," thereby leading him to assert, "… pride—where this is a real superiority of mind, pride will be always under good regulation," at which nonsense or inconsistency she doesn't dare laugh, but she does turn "away to hide a smile" (Part I, Chapter XI, Austen, *Pride and Prejudice*); or, finally, Mary Anne Evans'/George Eliot's *Middlemarch*—putting aside its socioeconomic and psychological realism, e.g., that of Lydgate's fall, he having too little reflected on his own traditional beliefs, among other things, thereby upending his medical ambitions and marital hopes—as funny and brilliantly metafictional, e.g., when a visual memory of John Locke's portrait, to which portrait witty commonsensical Celia refers, makes striking her criticism of Dorothea marrying, by her lights, ugly old Casaubon (Eliot, *Middlemarch*, 16), which bias later gets taken up at a metafictional level, when the narrator asks why one would favour young lively Dorothea's point of view over her husband Casaubon's, which question transitions the novel, as if making up for a lacuna, into his till-then-absent one (Eliot, 278).

73. Wallace, *This is Water: Some Thoughts, Delivered on a Significant Occasion, About Living a Compassionate Life*.

74. Wallace, "Track Three."

75. "[T]here is this existential loneliness in the real world. I don't know what you're thinking or what it's like inside you and you don't know what it's like inside me. In fiction I think we can leap over that wall itself [i.e., the one preventing this knowledge of other people] in a certain way" (Wallace, "The Salon Interview: David Foster Wallace.").

how they, finding community, better cope with loneliness and, especially, its effects, e.g., addiction. We, in turn, thereby, again and again, are to do the work and learn to see other people's stories as like our own—i.e., learn to better know others and empathize. In so doing, we are to learn to better cope with loneliness, as certain characters do in the story. That's the interpretation. Here's the recommendation based on it. The novel, for its entertainment and teaching, is good for head and heart. In this recommendation, notice the use of imaginings, i.e., that they're indirectly appealed to insofar as they support the entertainment and thereby the learning.

This last paragraph sketches a use of imaginings, a recommending one, in virtue of which they are of works of literary fiction. To see this, notice that such recommending is to literary practice like en passant capture is a move in chess, and the above sketch resembles describing such a move. After all, such recommending lies in literary practice. To see that it so lies, recall the practice. To do so, you might recall, first, that, as a matter of course, we often love to recommend, and we read to recommend or its opposite. You might also recall that, to get excited about reading a book, we read reviews, browse blurbs, and scan lists of nominated books. Finally, all of this, you might recall, we do for its own sake but, often, also to use the works in other ways, e.g., to participate in a book club, to grandstand, or to be a part of a grand literary institution, and so on.

A Sketch of Interpretation Uses

Turn now from recommending uses of imaginings to interpreting ones. To begin, let us, in line with the extrapolated approach,[76] distinguish two ordinary uses of "interpret." The first is "To expound the meaning of (something abstruse or mysterious); to render (words, writings, an author, etc.) clear or explicit; to elucidate; to explain…"[77] This definition allows sentences of the form, "Someone interpreted the obscurity for someone else." The second, roughly speaking, does not. It is "To make out the meaning of, explain to oneself."[78] This second definition, again roughly speaking, only allows sentences of the form, "One interpreted the obscurity for oneself." This "for whom" difference in usage is what I'm interested in here. To clarify, in line with §1.1.1, imagine a high school teacher assigning her students a book report on, say, Kafka's *The Trial*. She offers three writing prompts: On a religious interpretation of Kafka's *The Trial*, could K's purported crime be original sin and the trial, or process, be an allegory for the life of people in a fallen state? On a biographical interpretation of his novel, is *The Trial* a veiled history of his break with fiancée Felice Bauer, who has the same initial as a female character and love interest? On a historical interpretation, does the work capture the alienation of a people faced with opaque and oppressive governmental institutions? [79] Given these prompts, the students then write their papers. Now, we can say of these students writing their papers they may interpret the work in line with each of the two definitions. In line with the first, we can say they try to interpret *The Trial*, i.e., to render it clear or explicit on certain points, as if writing for their peers, and so get a good grade, and, the teacher, marking their papers, will, among other things, evaluate how well they've done it, when determining the grade. In line with the second, the students, preparing to write the paper, read Kafka's novel, try to interpret it, that is, to make out the meaning of the work for themselves.

Here we see that the two uses find different homes in describing different activities, one communicative, aiming to elucidate what's obscure to someone else, another contemplative, aiming to make sense of a work for oneself. To be sure, there's some overlap, e.g., reading out loud to exemplify interpreting. Also, the uses are inter-explainable, e.g.: in one direction, to make sense of it for oneself just is to elucidate it for others when the only other is oneself; in the other direction, to elucidate it for others just is to make sense of it for oneself when

76. Cf. *PI* §120.
77. *Interpret, v. 1. a.*
78. *Interpret, v. 1. b.*
79. Cf. Balint, *Kafka's Last Trial: The Case of a Literary Legacy*.

one does so publicly instead of privately. But, so far as I can see, neither limited overlap nor inter-explainability vitiates the distinction.

In sum, I'll use the term differently describing different activities. Specifically, I'll use it, primarily in the make-sense-to-ourselves way, to sketch some of our attempts to make sense of elements in a work—as we often do, by ourselves and by means of imaginings—while or after reading. I begin with those used while reading.

Interpreting with Imaginings while Reading

While reading to ourselves we often, if inadvertently, use imaginings to better interpret the work. Doing so sometimes consists in ordering information or else merely paying attention. To see this, consider an analogy, two examples, and three clarifications.

Here is the analogy. We sometimes read as if we are a trinity—as if, that is, we are a parent reading, the book's illustrator, and a child being read to who looks at the pictures. That is, the child makes better sense of the story the parent reads by means of the illustrator's pictures. For instance, while the parent reads the Berenstain Bears story *Trouble with Friends*[80], the child looks at the illustrator's pictures, which tend to capture the emotional high points of the scenes—such as the one in which Sister Bear and Lizzy Bruin, with open yelling mouths and slanted angry eyebrows, fight over a teacher's pointer—and, the child thereby better understands these scenes. That is, the child is more likely to answer more or less correctly when asked whether the cubs are fighting, why they're mad, and so on, instead of being wide-eyed dazed, saying nothing, whispering, "I don't know," or being easily distracted, and the like.

Here is the first example. Once I tried to figure out how, in Gaddis's *A Frolic of His Own*, the artwork *Cyclone Seven* looks, and how the puppy Spot is trapped in it, by picturing these things.[81] In particular, I seem to recall, while reading, first, visualizing, if inadvertently, a small dog behind long broad metal spikes which reach up high above the animal and then, second, to cope with further description of the artwork, adding to my imaginings, or changing parts thereof, to fit said description. That is, in this case, as I read the story, by imagining detailed story elements, I was enabled both to pay better attention to them and make better sense of them for myself.[82]

In this example, the many details would, without those imaginings, have hindered my attention and sense making. Imaginings can also help us pay attention in cases without such detail. The second example counts among these cases. Here it is. Once, reading Kafka's *The Metamorphosis*, I visualized an otherwise dull event—namely, Gregor Samsa, on his wall, his body covering the picture of a woman. Covering it, he's trying to keep his sister and mother from removing the picture. This occurs, unbeknownst to me at the time of reading, at a crucial juncture in Gregor's fortunes, after which they sharply decline, while those of his family rise.[83] That is, while reading this scene, which is hardly detailed, by visualizing something, I paid better attention to an important scene.

Finally, to preempt certain questions, consider three clarifications about imaginative accuracy, variation,

80. Berenstain and Berenstain, *The Berenstain Bears and the Trouble with Friends*.

81. Gaddis, *A Frolic of his Own*, 29 ff.

82. Here is a similar example. In one perplexingly populated scene in Eliot's *Middlemarch*, a group watches "old Featherstone's funeral from an upper window of the manor," the home of Dorothea and Casaubon (Eliot, *Middlemarch*, 325). The group consists in these two, her sister Celia with husband Sir James and mother-in-law Dowager Lady Chettam, her acquaintance Mrs. Cadwallader, and her uncle Mr. Brooks. They're high up, literally and figuratively, looking down on the funeral procession, the deceased a presumed but not actual benefactor to another main character, Fred Vincy, who proceeds alongside his family, relations of the deceased, and a pastor, all coming out of a church. Reading about all this, I found myself visualizing the upper room and a character or two looking out of the window, and then, to cope with complexity, I both added other characters and then, as if opening a new window on my mental screen, pictured the view out the window, which comprised the funeral service below. That is, while reading, by visualizing, I kept engaged and made sense of the scene.

83. Kafka, *The Metamorphosis*, 33–34.

and ordering. First, I'm describing a way in which we use imaginings while reading to better interpret elements of a work, and, so, I'm describing a use in virtue of which certain imaginings are experiences of a work. This presupposes that, while aiming to interpret a work, if we *rightly* picture elements of a scene to ourselves by sensorily imagining them, we experience the work in a certain way; and, to *rightly* picture by imagining in this way is, e.g., to visualize as if accurately illustrating a passage rightly read. By disanalogy, illustrations in a children's story are sometimes inaccurate. For example, on a page of Robert Munch's story *Marielou cass-cou*, illustrator Michael Martchenko drew many giant bandages, but the text describes only one.[84] If the text is fundamental, the story's illustration isn't accurate. Second, on this point, since embellishments needn't be inaccuracies, accurate imaginings may vary. By analogy, colouring the giant bandages beige or drawing the children carrying them as running, even though the words don't specify such colour or activity, embellishes but isn't inaccurate, and so another illustrator might equally accurately illustrate the bandages as pink and the children as walking. Third, illustrators, presumably, have usually first to understand the words before they illustrate. Conversely, we readers often imagine to understand and do not have first to understand in order to imagine. By analogy, it's like a child errantly doodling and then, when asked what is being drawn, looking at it, seeing a similarity to clouds, and then, without ever having meant to draw clouds, saying, "clouds." That is, we often absently imagine while reading and then, trying to interpret, use what we imagined.

To sum up, we have imaginings while or because of reading the work, and—when we use them to order or else pay attention to what we're reading and so to better interpret it—they are experiences of the work, so long as, e.g., we're accurately imagining what's rightly read.

Interpreting with Imaginings After Reading

So far I've sketched an interpretive use of imaginings had while reading. Let us turn to a use of them had after reading for the first time. It is that, often, we use them to see symbols or else merely to reflect on storylines. To bring this out, consider three sets of examples.

First, while *re*reading, we often use imaginings to recognize symbols, such as events that foreshadow or that introduce themes. One example is the train death early in *Anna Karenina* that, reading it again, I took to foreshadow Anna's own and to introduce such a fall's inevitability as a theme.[85] To so take it, I recalled the later event, which imaginings of Anna's impending death helped make memorable. Here, first-reading imaginings help second-reading symbol recognition, but elsewhere second-reading ones do. For example, take what happens to the dove, a symbol, at the beginning of Gaddis' *Carpenter's Gothic*.[86] Rereading it, I had the sick feeling that the novel is darker than I'd felt it was my first time through, struck as I was by the foreboding quality of the bird having been mangled and cruelly batted about like a shuttlecock by kids. This battered bird I imagined visually, which imagining presumably contributed to the foreboding feeling, that is, the emotional side of foreshadowing in this case. What's foreshadowed are actions in line with a major theme, which the bird introduces. The theme concerns the mangling of Christianity by unscrupulous or stupid people playing economic and political games.

For another set of examples, turn to uses of imaginings had after finishing a work but before reading it again. In particular, turn to uses merely for identifying symbols before we can interpret them, e.g., identify them as foreshadowing or as introducing themes. Again, take Kafka's *Metamorphosis*. Say we've been reading it and have come to the point where Gregor Samsa, who awoke as a large insect, flees his father throwing

84. Munch, *Marilou Cass-Cou*.
85. Tolstoy, *Anna Karenina*, 64.
86. Gaddis, *Carpenter's Gothic*, 1.

apples at him, one of which sticks to his back, which we imagine visually and which we find strikingly strange—i.e., want very much to explain.[87] That is, we read on to have it explained, much like we read on to find out what happens in the plot. Doing so, we find that the back apple rots. Then, against the rising and then golden Spring-like flourishing of Gregor's family, we're struck by the dark nightmarish quality of this back apple meanwhile rotting as he slowly dies. This apple, as we read, calls all the more strongly for interpretation. Now, the apple imaginings I recall hang together with memorable emotional responses, especially of dark, strange, bleak goings on, all of which come to mind when I try, rather fruitlessly, to understand the apple symbol. In this way, after reading Kafka's novella, I use the apple imagining to recall and mark off, as I assume I'm supposed to, the apple as a symbol. In short, imaginings inflame intentions to interpret.

Finally, here is the third set of examples. After reading, we often use imaginings merely to recall the work and neither to try nor successfully to interpret it. Put another way, we use them, as we look at pictures or movie clips, to relive works, to save them from oblivion, to savour them, and so on. For example, take Edith Wharton's *Age of Innocence*.[88] To recall its striking end, which occurs twenty-six years after the main storyline's end, I visualize the protagonist, Archer, sitting alone in a park below the rooms of his life's great romantic passion, Ellen Olenska, whom he is free to go up to and whom his son encourages him to go see—but he does not—instead, sitting there until a servant closes the doors of her balcony and turns off the light, at which time he arises and departs.[89]

3.3.2 Differences with Thought Experiments

To sum up, I've partly answered the question: How do we take imaginings had while or because of reading works of literary fiction to be experiences of it? My partial answer is that we often do so in virtue of using those imaginings, e.g., to interpret or recommend the work. To do so, recall, is like taking wooden figurines to be chess pieces by playing chess with them. Then, to fill in the answer, I described certain recommending and interpreting uses. That is, first, I described a way in which we appeal to a work's imaginings-supported cognitive and emotive features—thereby, recommending that work be or not be read—and, in so doing, taking the imaginings to be of the work. Second, I described two sorts of normal ways we so take imaginings by interpreting the work for ourselves. One occurs while reading and, in particular, while using the imaginings to order information, or else merely to pay attention. The other occurs after our first reading and, in particular, while using the imaginings to appreciate symbols in the work or else merely to reminisce about it.

These uses differ significantly from those in thought experiments. That is, they differ such that, if we try to explain how we learn from a story in a work of literary fiction—and in so doing either identify or regard it as a thought experiment—we risk losing our grip on it insofar as it is literature ordinarily read. My argument for this claim has two legs, one from recommending, the other from interpreting.

87. Kafka, *The Metamorphosis*, 37.
88. Wharton, *Age of Innocence*, 253–254.
89. In case this example seems far-fetched, consider two others. First, to recall a transition scene in *Infinite Jest*, I visualize, as if recalling a dream upon waking, the character Gately driving from the preceding scene's setting, a Boston half-way house, on his way to buy groceries, and passing, along the way, the hideout of an inept Canadian insurgent cell, the Antitois Brothers, at which point and place the following scene begins (Wallace, *Infinite Jest*, 475–480). Second, recalling the history Aeneas tells Dido in Book II of Virgil's *Aeneid*, easily to mind comes a sequence of visual imaginings, specifically, the Greeks entering Troy, Aeneas rallying troops, their putting on Greek armour as camouflage, Trojan archers mistakenly shooting them, Aeneas with household fleeing to ships, and his losing his wife (Virgil, *The Aeneid of Virgil: In the Verse Translations of John Dryden*).

Recommending Leg

To introduce the recommending leg, consider a recurring allusion to a famous preface of Joseph Conrad's[90] in Gaddis' fiction. In *A Frolic of His Own*, it is: "You remember Conrad describing his task, to make you feel, above all to make you see? and then he adds perhaps also that glimpse of truth for which you have forgotten to ask?"[91] Conrad describes a literary task. We cannot normally so describe Galileo's thought-experimental one. To explain why not, we might say that the great scientist tries to make us see—not merely to evoke. To say so isn't to deny that Galileo tries to evoke *in order to* make us see. Nor is it to deny that we do so—for example, when, to teach the famous Galilean thought experiment about falling bodies more interestingly, we replace the boring old stones with exciting cannonballs or add in Pisa's tilting tower. To say so, rather, is only to deny that normally we can describe the task as to make us feel and, in particular, to do so as an end and not merely as a means.

This difference between how we can normally describe Conrad's literary task and Galileo's thought experimental one is the general form of another difference, one between our uses of imaginings. To illustrate, many imaginings in *Heart of Darkness* help the book to evoke horror, and we can, as a matter of course, describe their doing so as Conrad's task, in addition to glimpsing truth, whereas we cannot normally describe Galileo's task as making us feel by getting us to imagine falling stones—much like we cannot normally describe an experimenter's task as making observations merely to evoke. To be sure, this isn't to deny that we can experiment, and thereby observe, with *ulterior motives*, such as merely to provoke or otherwise evoke; and, in this way, I can accommodate what's plausible in claims like Sorensen's that we can use experiments however we like, even to "work out a grudge against white rats."[92]

This difference, in turn, explains yet another one, which lies between recommending uses. That is, we can, as a matter of course, recommend a literary work like *Heart of Darkness* by appeal to its imaginings-supported emotions, such as horror, alongside its cognitive features; whereas, we cannot, normally, recommend a thought experiment like Galileo's Falling Bodies in part by appeal to what our imaginings evoke, e.g., to visualized cannonball coolness or falling stone homeliness. We cannot normally do so, as we can when recommending a work of literary fiction, because we cannot normally describe a thought experimenter's task as we often can that of a literary author's—that is, as to make us feel just as well as to see truths. Consider an alternate example of this way in which recommending a thought experiment differs. To teach external world skepticism, I might recommend the Brain in a Vat thought experiment, instead of only Descartes' demon one, on the grounds that its seemingly nearly-here technology will spark greater immediate interest and also shock at what, by the skeptic's lights, we do not know; however, what's evoked serves pedagogical ends and normally isn't to make students feel as any more than a means to a philosophical insight.

Consider a class of apparent counterexamples. The task of some thought experiments is to ask us what we should do in a given hypothetical scenario, and, in light of our choice, to show us that more matters to us than we think. We're asked, for instance, whether we'd plug into Nozick's experience machine and, in light of our presumed choice not to and a few arguments, we're to see that more matters to us than how things feel from the inside;[93] or, in an environmental ethics class, we're asked whether, were all sentient beings evacuated from Earth, we should blow it up for the enjoyment of all, and, in light of our presumed choice not to, we're to see, at least, that perhaps the Earth doesn't matter to us only in virtue of what's good for humans. Now, we can describe the task here as to make us feel, insofar as feeling figures in the outcomes, and on this basis we might

90. Conrad, *The Nigger of the 'Narcissus'*.
91. Gaddis, *A Frolic of his Own*, 318.
92. Sorensen, *Thought Experiments*, 205.
93. Nozick, *Anarchy, State, and Utopia*, 42–45.

recommend them; however, if fully described, we can only say that the feeling is a subsidiary outcome. That is, it is only a means to the overall outcome, e.g., that more matters than how things feel from the inside or that non-human nature may be intrinsically valuable.

To be sure, I allow that—in exceptional cases—we can so describe the task. We can do so when we regard the thought experiment as, e.g., an artwork or a mere curiosity or an interesting bit of class material. For example, I might recommend Gettier cases to a friend but not by appeal to it being a challenge to the thesis that knowledge is justified true belief. Rather, I might appeal to the amazement it evokes when we're asked whether So-and-So's true justified belief is knowledge and *find* that we're certain it isn't. Alternately, I might recommend Schrödinger's Cat but not for any light it sheds on how greatly a quantum phenomenon, superposition, differs from an ordinary macroscopic one, such as feline death, but for the sheer *Isn't-science-weird?* feeling evoked when we imagine the cat-vial-hammer-particle apparatus and are told the cat is, so far, neither dead nor alive. I'll explain why such cases are exceptional below.

Finally, with this difference between normal recommending uses in hand, suppose that we regard a story in a work of literary fiction as a thought experiment. Can we, as we are normally free to do—by appeal to one of the work's imaginings-supported emotions, alongside an insight—recommend that the work be or not be read? No, since, as we saw, we cannot normally so recommend a thought experiment. Now, recall that such recommending, as it were, glues imaginings to works; that is, such recommendation is one normal way in which we take imaginings had while or because of reading a work of literary fiction to be an experience of it. Regarding the story as a thought experiment, and being unable to so recommend, we're not free, as we normally are, to take, in this recommending way, such imaginings to be experiences of the work. That is, so regarding it, we cannot, as it were, stick imaginings to works with this recommending glue. To so regard it, then, is not to have one such glue. The same goes for identifying them. Doing so, then, puts at risk our grip on everyday literary practice and, in particular, on many imaginings in it. That is, thereby, we risk either (i) inventing a new and leaner use of imaginings for literature or else (ii) concerning ourselves with established but exceptional cases. By analogy, if we regard chess' bishops as checkers pieces, restricting their movement, we either invent a new game or else play an established but fringe one.

Interpreting Leg

That was the recommending leg of the argument. Having made it, I'll now give the interpretation one, which mirrors it, briefly, as follows.

Recall that, to help ourselves make sense of literary works, such as Eliot's *Middlemarch* or Kafka's *Metamorphosis*, we can, as a matter of course, use imaginings either while reading—to order information, or else merely to pay attention—or after reading—to appreciate symbols in the work, or else merely to reminisce about it. We cannot, by contrast, so use imaginings to interpret a thought experiment like Galileo's Falling Bodies or Thomson's Violinist—not unless it's as a means to seeing truth or the like. To elaborate these examples, first, if, while reading the Galilean thought experiment, we visualize a large stone freely falling near another which is attached to a small stone, we may thereby order and so keep track of what goes on in the hypothetical scenario, or else merely pay attention to it, but we haven't thereby made sense of it—not unless we see how it's a step toward seeing truth, i.e., that a certain Aristotelian thesis is false and a certain alternative true. Second, if, after having read Thomson's violinist thought experiment, you recall its scenario, visualizing the violin player attached to you, and, struck by the image, reflect that this connection is analogous to that between fetuses and pregnant women, then you recognize a symbol of sorts; however, you haven't made sense of that connection as part of the thought experiment, or even recalled it as such, unless you also

see it as a step toward seeing a truth, i.e., about whether a certain abortion argument's inference fails. Now, we cannot normally so interpret thought experiments, as we can works of literary fiction, because we cannot normally describe a thought experimenter's task as we can that of a literary reader. That is, we cannot describe it as making sense of matters, by so ordering or recognizing, with imaginings full stop—i.e., without allowing that doing so is a means to seeing truth or the like. To be sure, as above, in exceptional cases, we can so describe a thought experiment's task, e.g., if we regard it as an an artwork.

In light of this difference between normal interpreting uses, suppose, as above, that we regard a story in a work of literary fiction as a thought experiment. Can we, to interpret a work, as we are normally free to do, simply use imaginings to organize information or appreciate symbols, or else to pay attention or reminisce? No, since we cannot normally so interpret a thought experiment. As above, then, so regarding a story in such a work, we cannot, as we are normally free to, use such interpreting to glue those imaginings to the work. That is, since, recall, such interpreting is one normal way in which we take imaginings had while or because of reading a work of literary fiction to be an experience of it, and since regarding a story in such a work as a thought experiment constrains these normal ways of interpreting, to so regard it constrains our taking imaginings to be of the work in those ways. To so regard such a story, then, is to risk losing one's grip, since it's to remove a way of getting or keeping it. That is, to do so is to risk either inventing a leaner, more observation-like, use of imaginings or dealing in established but exceptional cases. The same goes if, instead of regarding one as the other, we identify the two.

Conclusion of Each Leg

In sum, if we regard a story in a work of literary fiction as a thought experiment, or identity one with the other, we're unable, as we normally are, to take an imagining, as we do in any of the described everyday ways of recommending or interpreting, to be an experience of the work. If we so identify or regard such stories, then, we risk losing our grip on the works encompassing them insofar as they're normally appreciated.

One may object that we risk nothing, since to so regard is only to focus on certain sorts of ordinary appreciation. That is, let it be that, by regarding a story in a work of literary fiction as a thought experiment, we cannot take an imagining had while or because of reading the work to be an experience of it as we normally do when, for example, recommending or interpreting, as described above. Even so, we are nevertheless free to take the imagining to be an experience of the work as we normally do, since we're able to so take the imagining as we do performing certain normal literary activities other than those described above. Specifically, we're able to do so performing those activities which use imaginings solely as a means to seeing truth or the like. To so regard is to draw our attention to such activities. Doing so, then, hardly risks our grip on literary works. My replies are twofold and modest too. The first is that, without a better description of this strictly epistemic use, its existence is presupposed for the objection's sake, and, consequently, the objection merely amounts to a vague worry. This downgrades the objection but doesn't erase it, since we might investigate and find that we can give the description. To further downgrade it, here is my second reply, an in-principle one. It is that, so far, it's not clear that we can focus, as it were, on certain moves in the literary game without turning from the whole. By analogy, we can regard bishops in chess as pieces in checkers, focusing on similar diagonal ways both move, but, since the bishops can also move differently, e.g., farther and backwards, to so regard them is also to turn from chess, since it, unlike checkers, grants further—even if unexercised—freedom to its pieces. Similarly, suppose there exists the presupposed everyday literary use of imaginings that aims solely at seeing truth or the like and that, to draw attention to this use, we regard the related story as a thought experiment. In so doing, we might risk prohibiting potential everyday uses of the

stories in virtue of which they count as belonging to a given work of literary fiction. For example, an evoking use of imaginings in action scenes is not obviously separable from their use as recollection aids or information organizers or something more akin to observation-like imaginings in thought experiments. That is, it's not obvious such uses aren't all of a piece such that to turn from some is to turn from all.

Finally, the conclusion of the above two legs serves three purposes. First, it's a main plank in my criticism, in §3.2.2, of Elgin's explanation of how we learn from stories in works of literary fiction. It is also part of a plank in §3.2.1's second criticism of Davies' argument. That is, the group of differences between literary uses of imaginings and those in thought experiments is one of the three that helps to shift the burden of proof back onto his identity claim. Third, in virtue of this group, the conclusion also challenges accounts like Davies', e.g., Davenport's, that rely upon an identity claim.

Before moving on to these other differences, as promised, I'll explain exceptional cases. Also as promised, to develop my criticism of Elgin's account, but also to illustrate the above imagination difference, I'll offer three case studies.

Regarding Thought Experiments as Artworks

To clarify my claim that we cannot *normally* describe the task of a thought experimenter as to evoke, as an end in itself, I pointed out a contrast case, i.e., an exceptional one, in which we can so describe it—namely, that of regarding a thought experiment as art, or else as a curiosity or pedagogical aid. I promised further clarification here. To follow through, I will explain why this case is exceptional. It is, in short, because we do not recognize thought experiments to be artworks or curiosities or such aids—at least not as, seeing faces and saying names, we recognize acquaintances. To bring this out, I'll compare this regarding to that of literary works. As a preview, the main difference will be that, whereas we can regard a thought experiment as an artwork, we can't regard a work of literary fiction as one, for that is what it is.

I'll elaborate a little and then some more. We can describe a thought experimenter's task as to evoke with imaginings, if we regard the task's result as an artwork; but we're not enabled, by regarding a literary work as art, to so describe it, since, apart from the fact that we're already able, that's what it is. Furthermore, we can recommend a thought experiment, based on emotional features which imaginings underlie, if we regard it as an artwork; but, for those same reasons, we're not enabled to do so for literary works by regarding them as art. Finally, we can interpret a thought experiment, using imaginings to order information or recognize symbols, if we regard it as an artwork, but, again for those reasons, we're not enabled to do the same w.r.t. literary works. Now, to elaborate some more, I'll ask why we can't do so in one case and, in that light, why we can in the other.

Let us ask, first: Why can't we ordinarily regard the literary work as art? An analogy will help bring out the answer. Normally, we cannot—looking at the punctuation mark ":" and saying, "the colon is the colon"—regard it as a colon; rather, we are still only looking at the colon. Neither can we so regard it merely looking and saying, "the colon is a punctuation mark." In special circumstances, however, we can. We can, for example, after regarding the colon in the emoticon ":)" as eyes, regard it as a colon, looking at it and calling it so. Normally, that is, the expression "to regard as" is like "to see otherwise"—insofar as it has a logical form akin to *S regards X as Y* only if $X \neq Y$ or akin to *S regards X as* $\in$ *Y* only if $X \notin Y$—and, consequently, as a matter of course we cannot apply it to looking at and recognizing something. In special circumstances, by contrast, we can so apply it, e.g., say, "we regard X as we normally do," after talking about regarding X as Y where $X \neq Y$. You may, now, be tempted to analyze "look at and recognize X" in terms of "regard X as we normally do" or "regard X as familiar," thereby making more tractable the mess of terms we've amassed, but please hold off on doing so, since this reduction threatens to paper over the difference I've been trying to bring out between what

we normally call "looking and recognizing" and "regarding as." Now, with this difference in view, here is an explanation analogous to the promised one. Normally, we cannot—having a work of literary fiction in plain view and saying, "it's a work of literary fiction," or, "it's art"—regard it as one or the other, respectively, since as a matter of course we already recognize that it is one or the other. To be sure, in special circumstances, we can so regard it, e.g., if I, failing to recall that Borges' famous short story *Pierre Menard, Author of the* Quixote is the short fiction readers normally know it to be, call it an essay, as if it were an ordinary academic journal article, you could correct me saying, "recall that it's a work of literary fiction, a short story in essay form" or, "it's an artwork," and thereby I may be said to *regard it as one.* In light of this similar explanation, finally, here is the promised one. We normally cannot regard a literary work as an artwork because, as a matter of course, we already recognize that it is one. For example, I may recommend Borges' story as charming, and not only as interesting for its ideas, touched on below, about authorship; and, in support, I may point out an imaginings-underwritten happy footnote about book burning.[94] In giving and receiving such a recommendation, or interpreting such a passage, normally, we recognize a work of literary fiction—and so we cannot regard that work as one. To be sure, in special circumstances, one can so regard it, e.g., if we've been imagining that the story really was published as an academic article and now return, so to speak, "to regarding it as we normally do, as the artwork it is."

Second, given that we cannot normally so regard these works, how is it that we can so regard thought experiments? Well, in part, because we do not normally recognize art, or a curiosity, when we encounter a thought experiment—not, by contrast, as we recognize something experiment-like. To illustrate, it takes some special effort or circumstance to both pay close attention to emotional features in a thought experiment—such as how exactly its scenario is wacky, entertaining, or mind-bending, or how its expression is beautiful, elegant, or even lapidary—and ignore various experiment-like properties in it, such as how what's imagined bears on our knowledge or understanding. By contrast, normally we cannot *regard* thought experiments as, e.g., a means of gaining knowledge or understanding—since as a matter of course this is part of what we see when we recognize them. In sum, we can as a matter of course regard them as artworks, or else curiosities or pedagogical aids, since, although similar in various respects, we do not normally recognize them as such upon encountering them.

Developing the Criticism of Elgin's Account: Three Case Studies

Case Study I We may recommend what is perhaps Borges' most famous short story, "Pierre Menard, Author of the *Quixote*," by saying that, reading it, we learn a real-world technique for interpreting stories. In the story, an academic of sorts aims to account, among other things, for the efforts of one Pierre Menard who, without becoming Cervantes, tries to write *Don Quixote.* The lines Menard writes must comprise the same words in the same order, but, as the academic argues, their meanings change with the change in author and, in this case, for the better. To illustrate, Borges' academic makes the following, perhaps unwittingly funny, comparison:

> It is a revelation to compare the *Don Quixote* of Pierre Menard with that of Miguel de Cervantes. Cervantes, for example, wrote the following (Part I, Chapter IX):
>
> … in truth, whose mother is history, rival of time, depository of deeds, witness of the past, exemplar and advisor to the present, and the future's counselor.
>
> This catalogue of attributes, written in the seventeenth century, and written by the "ingenious layman" Miguel de Cervantes, is mere rhetorical praise of history. Menard, on the other hand, writes:

94. I.e.: "I recall his square-ruled notebooks, his black crossings-out, his peculiar typographical symbols, and his insect-like handwriting. In the evenings he liked to go out for walks on the outskirts of Nîimes; he would often carry along a notebook and make a cheery bonfire" (Borges, *Collected Fictions: Jorge Luis Borges*, 95).

> … in truth, whose mother is history, rival of time, depository of deeds, witness of the past, exemplar and advisor to the present, and the future's counselor.
>
> History, the *mother* of truth!—the idea is staggering. Menard, as contemporary of William James, defines history not as *delving into* reality but as the very *fount* of reality. Historical truth, for Menard, is not "what happened"; it is what we *believe* happened.[95]

After the academic reflects on this example, the story ends so:

> Menard has (perhaps unwittingly) enriched the slow and rudimentary art of reading by means of a new technique—the technique of deliberate anachronism and fallacious attribution. That technique, requiring infinite patience and concentration, encourages us to read the *Odyssey* as though it came after the *Æinid*… This technique fills the calmest books with adventure…[96]

That is, Borges' academic ends the article recommending an interpretive technique—to carefully read, e.g., the *Odyssey* attributing each line to, e.g., Virgil instead of Homer, thereby changing the lines' meanings and making it more exciting. Again, Borges' academic, perhaps unwittingly, produces smiles or a raised eyebrow, with clumsy or curious diction: As if the *Odyssey* were otherwise among "the calmest books"! As if what requires "infinite patience and concentration" could fill anything with adventure! Still, despite this humour or irony about the technique, the idea that one might so apply it to the *Odyssey* has its charm.

To bring it out, we imagine reading the epic assuming Homer wrote it after Virgil the *Æinid*. Doing so, the literary origins reverse themselves, and we may see Homer as subversive. Take Virgil's story of Troy's fall. Reading it as we normally do, we see Homer's story of the city's fall from another point of view, that of the hero Æinias. For example, we see Odysseus, in devising the Trojan Horse, not as a resourceful hero but as a scheming villain, and the city's fall begins a voyage not back home but to found a new one, namely, Rome. Now, apply the technique, instead of reading as we normally do. If the *Æinid* were written first, we'd recognize the *Æinid*'s story of the city's fall in the *Odyssey*. The origins would then appear reversed, since Homer would seem to draw on Virgil. Then we could ask, if *Homer* turned it on its head, why? To subvert Roman authority, to write a rival chest-thumping nationalist story, one for the Greeks and not the Romans, one about returning home to Hellenism and not one of Rome's founding?

In sum, the academic, among other things, illustrates a technique, that of interpreting a work with "deliberate anachronism and fallacious attribution," and encourages its use; and, with this illustration and encouragement in mind, we may recommend Borges' short story saying that it humorously both teaches one a reading technique and encourages its application.

Consider four worries about this recommendation. First, in light of the humour or irony that Borges directs at the academic's way of treating literature, isn't the technique not recommended but used to lampoon or satirize such academic ways? To ease the worry, we may interpret the humour or irony another way, as a means to get readers through an otherwise difficult story, even if upon close scrutiny we'd agree that it's a lampoon or satire, since we're dealing with everyday reading; and, we may append to our recommendation that we're so interpreting it. This leads to another worry. It's that, since the application of the technique which underlies the recommendation relies upon some possibly uncommon knowledge, e.g., about the *Æinid*, I do not deal, as I should, with reading the short story as we ordinarily do. There's no need to worry here, since by reading the story one need only find the idea that someone write *Don Quixote* again intriguing and want to apply the idea in light of reading the *Odyssey* as if the epic were written in some other context. Third, how could we possibly learn a reading technique from a fictional story about it? We might appeal to Elgin's theory. To do so, we could, regarding the story as a thought experiment, point out that the comparison, which the

95. Borges, *Collected Fictions: Jorge Luis Borges*, 94.
96. Borges, 95.

academic makes and Borges contrives, between what Menard writes and Cervantes' syntactically identical expression, quoted above, exemplifies the concept of "deliberate anachronism and fallacious attribution." Thereby, we gain "epistemic access" to possibilities of interpretation in stories, such as that above of reading Homer's *Odyssey*. Finally, fourth, how are we granted such access if we're not cued to it? By analogy, we can't apply the phrase, "big brother is watching," to cases of widespread, systematic and totalitarian surveillance, if it won't come to mind in appropriate circumstances. Well, supplementing Elgin's account a little, we may add that Borges, by means of the academic, cues us to such "access" by encouraging the technique's use.

At this point, defending the recommendation, we have applied Elgin's theory to our use of a story in a work of literary fiction to explain how we could learn from it. Theory applied, is the use a normal everyday one? On the whole, no, for certain everyday literary features cannot figure in it. To begin, by rough analogy, to explain Swift's *Modest Proposal* as mere polemic hardly explains the satire. Next, by example, so explaining, we leave out the imagining-underlying humour or cheeriness and, more to the point, we cannot add these to the story's use. To see why not, first, note the lighthearted or funny, and imaginings-underwritten, footnote cited above about Menard destroying his writing material in a cheery fire. These evocations, moreover, hang together with headier aims. For example, both the burning of Menard's writing materials and his death diminish how well one can find out what he meant by any given line in his *Don Quixote* and so renders the work more ambiguous, that is, as the academic would have it, richer—since "ambiguity is richness"[97]—and, more to the point, riper for the application of the very technique he recommends. Now, we cannot add this evocative use of imaginings to the story's use, as we regard it when applying Elgin's theory. This is because, doing so, we could not, as we're normally free to, take the emotions to be ends in themselves instead of merely means to seeing truth or, in this case, understanding, and knowing we understand, the reading technique. We could not because we're regarding our use of the work's story as a thought experiment, which has no such role for these emotions, unless we see it as something else, e.g., an artwork. Applying Elgin's theory, then, we risk losing our grip on Borges' story.

We can give similar accounts of some classic and contemporary dystopian works, such as *1984* and Naomi Alderman's *The Power*. Schematically, we, having interpreted some of the work, recommend reading it by appeal to what one would learn, and, to explain how we so learn, we apply Elgin's theory and, in so doing, regard a story in the work as a thought experiment; next, we point out that, in so explaining, we restrict the freedom we normally have to use imaginings either to evoke, order, or recognize symbols as ends in themselves—and thereby take them to be experiences of the work. In the following two case studies, I'll consider such symbol recognition and information ordering.

Case Study II To illustrate this recognition, we might recommend *1984* to understand contemporary political discourse. We may do so, first, recalling, by means of imaginings, Winston's colleague Syme in the Ministry of Truth—i.e., the intelligent, elitist one with a passion for writing the Newspeak dictionary, which writing, by reducing the language's word count, reduces its expressive power and even "thought, as we understand it now;"[98] and, second, we may do so interpreting Syme's job as a symbol of language-based thought control. We may then be asked how we so learn. To explain, as Elgin does, we may, regarding Syme's story in the novel as a thought experiment, point out that the contrived situation in which this character acts exemplifies how limiting language limits thought, adding perhaps that we use imaginings to grasp the concept exemplified, and that this—plus perhaps knowledge of Orwell's political aims from such a work as "Politics

97. Borges, *Collected Fictions: Jorge Luis Borges*, 94.
98. Orwell, *Animal Farm & 1984*, 136.

and the English Language"[99]—gives us not only "epistemic access" but motivation to use it and so to see how watering down real-world political discourse may limit thought. Finally, we point out that, in so explaining, we restrict our normal usage of imaginings of that character in the work—e.g., preclude merely using them to recall or to evoke Syme's tragic character as an end in itself. This tragedy lies in his fate:

> One of these days, thought Winston with sudden deep conviction, Syme will be vaporized. He is too intelligent. He sees too clearly and speaks too plainly. The Party does not like such people. One day he will disappear. It is written in his face.[100]

In sum, we risk losing our grip on Orwell's novel as literature normally read. That said, I'd like to stress, in light of the novel's political importance, that all this is neither to deny that we can use the stories other than in normal literary ways nor to deny that we should—nor is it to deny that we frequently do so learn because of reading the book aiming to learn such things from it—nor is it to deny that Elgin's account explains such learning. On why the risk nevertheless matters, recall §3.2, on literature ordinarily read.

Case Study III To illustrate the information ordering, consider *The Power*, a dystopian novel composed of two stories. The main story, set in our times, omnisciently chronicles the rises and falls of its four protagonists in light of the novel's primary premise—that young women, mostly, discover that they command great electric power. This main story is a historical narrative written in the other, a limited third-person point of view, framing story—which is set after the main story's events, the subsequent collapse of patriarchal civilization, and the rise of a matriarchal one. We may recommend it by appeal to something one can learn about power and gender. To begin doing so, we may use imaginings to recall and order information—e.g., to call to mind various similar "perspective switches" and place them side by side so as to bring out a pattern that spans scenes in both storylines. Some examples: by visualizing an ancient artifact of a woman electrifying a man's genitals to control erections, described in one of the main story's historical interludes, we recall a scene in which there is not female but male genital mutilation to control sexual behaviour; by visualizing the sexily clothed young man forced to lick up liquor he spilled, glass shards cutting and entering his tongue, we recall a scene in which there is not the objectification and degradation of females by powerful males but of males by powerful females; by visualizing the superhero-like exploits of the preeminently powerful protagonist Roxy, we recall scenes in which it is not an active male hero kicking villainous ass and saving passive women but a female hero male ones; by imaginings of the protagonist Mother Eve converting through online video clips, we recall a scene in which she uses not the capitalized pronoun "He" to refer to a male deity but "She" to a female one, one like those newly powerful women she wishes to influence; by vague visualizations of a city far from the male protagonist Tunde's surroundings, I recall a scene in which it is not a woman's intellectual property stolen with impunity for career advancement but a man's; finally, by a vague visualization of framing-story character Neil's university building, I recall a scene in which that character speaks not of a Gender Studies department and its research as being taken lightly but of a Men's Studies one and its research being so taken. To describe this pattern, we might say that the author, Alderman, in many scenes, switches certain characteristics of, or relations between, men and women to reveal how power determines those characteristics and relations. Having so used imaginings to order the information, we may appeal to this ordering both to explain what it is that we can learn reading the book and thereby to recommend reading it. We may then be asked how we so learn, and, to explain, we may appeal to Elgin's theory. To do so, we, first, regard these scenes as a series of hypotheticals in a thought experiment. That is, we, as it were, take the novel to ask what would happen were

99. Orwell, "Politics and the English Language."
100. Orwell, *Animal Farm & 1984*, 137.

women to gain such electric powers and, having us imagine what would, answer that it would turn patriarchy to matriarchy. Second, we may point out that these scenes are contrived to exemplify how, roughly, power differences determine gender differences. If we're unconvinced that thereby we gain "epistemic access" to the world, we might add to Elgin's account that as we read the scenes, which repeatedly exemplify similar ways power differences do so, we develop certain habits of mind which we carry with us, as it were, outside the work of literary fiction—cueing us to notice relevant aspects of power in the world. For example, I found myself, for some time after having finished the novel, inclined in my daily life to switch perspectives, or try to, as in the above scenes. But, finally, by my lights, in so explaining, we risk inadvertently turning from what we want to explain, i.e., from the novel as ordinarily read—for, in so explaining, we restrict what imaginings-supported features of the work can be for and, by extension, the ways certain imaginings count as being experiences of it. For instance, our imaginings support action sequences in which Roxy righteously wields some serious voltage to bring down both henchmen and their villainous boss, or so it seems before certain plot reversals. To require that this entertainment only serve to improve our understanding of power and gender, and not this also as an end in itself, is to risk losing one's grip on the novel as literature normally read.

To be clear, I do not deny that, to *interpret* a work like *The Power*, we should regard it, or a part of it, as a thought experiment. For here we're not engaged in literary criticism. To illustrate, consider:

> World-building quibbles aside, it's difficult to bear [*The Power's*] conclusion that the horrors of our times are inevitable and inescapable: that there will always be abuses of power, that the arc of the universe doesn't bend toward justice so much as inscribe a circle away from it, that if our world were destroyed and rebuilt with women in charge it would look exactly as it does with men in charge. The tension between thought experiment and gripping realism is tricky to navigate, and it left me wanting to argue, without quite knowing what the book's position ultimately was. To show up the double standards between men and women? To enact what Sophia MacDougall wrote about in her essay "The Rape of James Bond" and show horrific instances of equal opportunity rape? These are both done to tremendous effect—but I found myself questioning the aim.[101]

In short, for critic Amal El-Mohtar, this novel/thought-experiment ultimately fails because of its unbearable or questionable, albeit unclear, outcome. We might disagree. For instance, we might say the novel is instead about how imbalances of power lead to them or simply about perceiving such imbalances behind sexism. Similarly, and more to the point, we might deny that we should read the novel as a thought experiment. To explain why, we might say that to do so requires something that the work need not have to succeed, namely, a clear overarching outcome. Now, to so disagree is to engage in literary criticism. What's at stake, among other things, is how we should appreciate the novel. Contrarily, to apply Elgin's theory, as we do, isn't. Rather, it's to stand outside the practice and explain how, appreciating the novel, we could possibly learn from it. The upshot is that, in criticizing such theories' application, I haven't denied that one should, or should not, interpret a work as a thought experiment.

3.4 Outcome Differences & Overall Differences

Last section, I explained important differences between certain normal uses of imaginings in works of literary fiction and those in thought experiments. That is, first, I gave a partial account of how we ordinarily take imaginings when appreciating works of literary fiction to be experiences of those works. On this account, we do so using the imaginings ultimately to recommend and interpret; to do so, among other things, we use them for certain literary effects, such as symbol recognition, for the good of head and heart, such as entertainment and learning to cope with loneliness, and for story recollection—all of which are literary goods in themselves.

101. El-Mohtar, "March's Book Club Pick: 'The Power', by Naomi Alderman."

Second, I described certain ways in which these uses of imaginings differ from how we would take them were we either, like Elgin, regarding the work's stories as experiments in thought or, like Davies, identifying the ones with the others. If successful, I thereby offer one challenge to both Elgin's and Davies' explanations, introduced in §3.2.1 and §3.2.2—aiming, ultimately, to clear away a problem about how we learn from stories in those works, in line with Chapter 1. Now, to this same end, I'll point out two other differences. They do not concern imaginings so much as outcomes, specifically interpretive freedom, and an overall property, i.e., complexity.

I do not claim that these other differences are the only other relevant ones. Other possible ones worth exploring include these four. First, typical thought experiment hypotheticals may resemble standard literary narratives only in broad outline. After all, we can ask: Are they stories with beginnings that establish, among other things, characters, situations, and themes? With middles that, among other things, develop characters, raise the action, and enrich themes? And with ends that, among other things, settle the fortunes of characters, fill plot holes, and polish up themes? In this light, is the notion of "fiction" used in accounts like Davies' and Elgin's too thin? If so, does it render the following claim of Elgin's false?

> Thought experiments can be construed as tightly constrained, highly focused, minimalist fictions, like some of the works of Jorge Luis Borges. If the minimalist stories of Borges are genuine fictions, there seems no reason to deny that thought experiments are too.[102]

Second, the figures in typical thought experiment hypotheticals may most resemble stock or flat characters in literary novels, or placeholders for ideas, not central round or dynamic ones. To illustrate, even if we see in Raskolnikov, from *Crime and Punishment*, a character type, that of the "underground man," and an enactment of socialist/atheist ideology,[103] we nevertheless treat descriptions of him like those of an old friend with terrible misguided morals, and we feel relieved or asphyxiated as his fortunes rise and fall and rise. Third, whereas imaginings in a typical thought experiment may be thought of as data or information, we might only be permitted to think of those in standard literary narratives as organizing or shedding light on the story or, as it were, the data. Fourth and finally, like the notes of Für Elise, the words of a literary work are to be treated as aesthetic objects seemingly unlike those expressing a typical thought experiment. For instance, if you rewrite *Pride and Prejudice* and end up with the novel *Pride and Prejudice and Sea Monsters*, you've written a new story, a piece of fan fiction perhaps, which is of lower aesthetic quality insofar as it's a sort of copy. Alternately, if, in light of James Joyce's nachlass, you change certain words in *Ulysses*, as if restoring a painting, you've improved its aesthetic quality insofar as it's better in line with original authorial intention. But the words we read to understand a thought experiment we seem permitted not to treat as having such aesthetic qualities. For instance, if we consult Galileo's *Two New Sciences* and determine that the falling bodies thought experiment involves stones, instead of the cannonballs it's so often taken have in it, then we will not bemoan slights against authenticity, as we might freely made changes to Beethoven's score or a jazzing up of Austen's diction.

3.4.1 An Outcome Difference: Interpretive Freedom

Here is an explanation of the outcome difference on which I'll focus—overstated for clarity. We can interpret stories in works of literary fiction more freely than we can make sense of a thought experiment's hypothetical scenario. After all, I'm free to read and learn from such a story however I like, but I cannot draw from a thought experiment any conclusion but the one or ones proper to it. This explanation overstates the difference in both directions, since story interpretation has less freedom and thought experiment more. To explain, take

102. Elgin, "Fiction as Thought Experiment," 230.
103. Cf. Frank's view below, in §3.4.2.

each direction in turn. First, Margaret Atwood has said, in an afterward to *The Handmaid's Tale*, that the novel responds to the claim, "It couldn't happen here"; and, she adds, she drew every fictional situation from an actual one.[104] We need not think Atwood's intent determines the book's proper interpretation, and we need not require knowing it to appreciate her novel, but, once we know such an intent, we normally take it to be significant, e.g., are guided by it or rebel against it. Either way, once known, it normally constrains how we interpret—much like Galileo's intent does his falling bodies thought experiment. In this way, I'm not that free to read and learn from such a story however I like. Second, we do, sometimes as I'll touch on below, describe one clock-in-a-box thought experiment without specifying either Einstein's or Bohr's outcome, or the disjunction of them, as *the* outcome. Also, we do sometimes treat Thomson's violinist thought experiment as we may a metaphor, namely, like an open-ended invitation to compare as opposed to an assertion of the form "A is to B as C is to D," as Aristotle had it.[105] Doing so, we vary the conditions. For example, we replace the violinist with an old dying dog or Nelson Mandela or Donald Trump; or else we add that the Society of Music Lovers will enrich Oxfam so long as you're plugged in; or else we add that the violinist awakes and looks you in the face and pleads with you not to unplug yourself; and then we think through what we're morally obliged to do. So we do have some freedom to draw conclusions from thought experiments.

To see that a difference remains, notice that, often, as a matter of course, we distinguish named thought experiments by their outcomes as if they were arguments, but we do not, in quite the same way, distinguish works of literary fiction, or the stories in them. For example, you have read and understood Galileo's Falling Bodies Thought Experiment only if you recognized that it aims to destroy the Aristotelian view that, roughly, bodies freely fall at a rate proportional to their weight, and to establish another, that they fall at an equal rate regardless of weight. This is so even if, carrying it out for yourself, you varied the kind of objects that fall and how they are attached and then thought through the consequences. By contrast, you may unquestionably have read and understood Atwood's *Handmaid's Tale* even if you did not hear that her intent was to show that "it could happen here," or even if you did hear but, resisting matters, nevertheless read the work as mere dystopia. This is so even though the novel has an intended outcome like many typical named thought experiments.

Again, the difference is that we can interpret stories in works of literary fiction more freely than we can make sense of a thought experiment's hypothetical scenario. The explanation, now not overstated, is this: as a matter of course, we need not take authorial intention to determine what we should learn from a story in a work of literary fiction; whereas, on pain of misunderstanding a named thought experiment, we normally do have to take its intended outcome to be essential to it. To fill in this explanation, I'll raise an objection and reply to it.

Objection: No Difference Exists

There's no such difference, the objection runs, since thought experiments never have their outcomes essentially; rather, they have only those historical outcomes that the Galileos and the Newtons and the Einsteins had carrying them out—much like experiments do, i.e., have results that aren't essential. Rather, they're just the ones their originators happened to derive.

To supplement this objection, and explain its importance to me, consider two of Elgin's arguments that, in effect, thought experiment outcomes are *invariably* accidental. We see it in an argument, from disagreement among the authors of the EPR thought experiment about what its hypothetical shows, that "thought

104. Atwood, "Afterward."
105. Cf. Hagberg, "Metaphor."

experiments require interpretation" and that "Sometimes interpretations diverge."[106] We also see it in this argument:

> [A thought experiment] is subject to… reinterpretation if the background assumptions change. Schrödinger's cat, originally introduced to criticize the Copenhagen interpretation, now appears in every interpretation of quantum mechanics, each offering a different account of the poor beast's state.[107]

In these arguments, Elgin assumes expressions like "EPR thought experiment" and "Schrödinger's cat" are kind terms and not names, which they often are but also often are not—and named thought experiments typically have their outcomes essentially. Using the terms as names, we can felicitously ask, for example, whether the EPR thought experiment even has an outcome, in light of the disagreement, or what the real, or original, outcome of Schrödinger's Cat thought experiment is, in light of the differing accounts of the cat's fate. That is, in short, the arguments unjustly favour a way of thinking about thought experiments, and an objection, based upon them, that thought experiment outcomes are necessarily accidental, or vary with background, cannot, as it stands, succeed.

By the way, Elgin allows for something like this difference. She casts it in terms of typical thought experiments not bearing multiple correct interpretations, like works of fiction, but instead being ideally univocal, given a set of background assumptions.[108] I'll come back to this below, in §3.4.2.

Now, to shed some light on this response, consider an analogy. Michael Bishop influentially objects to John Norton's account of thought experiments. The objection models thought experiments on experiments in a special way that precludes our modelling them on arguments, and this is to unjustly privilege a way of thinking about them. That's the gist of the analogy. Details now follow.

Reply: Two Models of Thought Experiments

Originally, John Norton restricted his account of thought experiments to those in modern physics and to specifying two necessary conditions on the kind of arguments they are. This was all he needed then. The two conditions are the following: an argument is a thought experiment only if it "(i) posit[s] hypothetical or counterfactual states of affairs, and (ii) invoke[s] particulars irrelevant to the generality of the [argument's own] conclusion."[109] For example, to satisfy (i), we might posit in the premises of an argument a hypothetical state of affairs by saying, "suppose that there's this speedy elevator with an industrious scientist inside and…" To satisfy (ii), well, we've already invoked particulars, the elevator and such, so we need now only make those particulars irrelevant to the generality of the conclusion; to do this we could make the conclusion about all such particulars—e.g., not just about that elevator and scientist and so on but about all masses and observers and so on. Later, he recasts these conditions. He defends an "… account of thought experiments as ordinary argumentation that is disguised in a vivid pictorial or narrative form."[110] The "core thesis" of his account is still that "Thought experiments are arguments."[111] The two necessary conditions, it seems, are eased into one on the form an argument must take to be a thought experiment, and he expands his account to be of, not just thought experiments in modern physics, but those in the sciences, and he expects it to hold in other places as well.[112] In any event, the ones he concerns himself with tend to be those in the sciences, and in particular

106. Elgin, "Fiction as Thought Experiment," 230.
107. Elgin, 230.
108. Elgin, 239.
109. Norton, "Thought Experiments in Einstein's Work," 129.
110. Norton, "Why Thought Experiments do Not Transcend Empiricism," 45.
111. Norton, 49.
112. Norton, 44.

those in the natural sciences that yield "contingent knowledge of the natural world."[113] We saw above, in §1.2.2.1's extrapolation, how he gets to this recast position, which he develops and defends. But let us move on.

Michael Bishop objects to such argument views of thought experiments, with Norton's version as the representative case. The objection is important in the literature because many philosophers have built on it.[114] At first pass, the objection is that the argument view cannot make sense of certain disagreements about thought experiment outcomes. For a second pass, first, Bishop takes the following representative case of disagreement about a thought experiment outcome from the history of science.[115] Einstein presented the clock in a box thought experiment at the 1930 Solvay Conference on magnetism. The thought experiment was aimed at the Heisenberg Uncertainty Principle. This principle, whatever else it does, expresses a limit on how accurately you can measure conjugate pairs, such as time and energy. What Einstein did to challenge this principle was to have his audience imagine a box. In this box is a clock and a bunch of photons. The box is weighed. At a later time, a door in the box opens and out goes a photon. The time the photon left is noted. The box is weighed again. Now, roughly speaking, from the difference in weight and the time the photon left, we can in principle calculate that photon's energy to any degree of accuracy. So there's no limit. This contradicts the Heisenberg Uncertainty Principle. For this reason, thought Einstein, we should give it up. Bohr, then, the history goes, who was at the conference, comes back the next day with a more detailed picture. He's added to his picture various instruments, various procedures and, crucially, relativistic spacetime. (Yes, he added Einstein's own theory!) Using this new picture, Bohr argues that it's impossible in principle to measure the energy and time to any arbitrary degree of accuracy and, stunningly, that the limit is indeed that specified by the Heisenberg Uncertainty Principle. The physics community agreed with Bohr that Einstein hadn't refuted the principle. Now, second, Bishop goes on to argue that this presents a counterexample to the argument view.[116] The bare bones of the argument have an abstract and a concrete part. The abstract part is this. Thought experiments are repeatable. That is, every thought experiment type can have multiple tokens. Now, on the argument view, thought experiments are arguments, and Bishop takes this to mean that each thought experiment type is identical to an argument type. So, for Bishop, it follows that, if two things are tokens of the one thought experiment type, they are both tokens of one argument type. Now for the concrete part. The clock in the box thought experiment was repeated; that is, this thought experiment type had two tokens, Einstein's version and Bohr's. So—on the argument view—Einstein's version and Bohr's are two tokens of the same argument type. But they're not, for they had different conclusions. Einstein's conclusion was that Heisenberg's principle is false, Bohr's the opposite. So, infers Bishop, the argument view is false. In short, on the argument view, some thought experiment types are identical to different argument types, and so, on pain of violating the transitivity of identity, the view is false.

Now Norton's no slouch. He responds that Einstein's and Bohr's versions are two different thought experiment types.[117] Since Einstein used classical spacetime whereas Bohr used relativistic spacetime, the hypotheticals are different, and, consequently, so too are their versions. That is, different thought experiment types are identical to different argument types. Transitivity of identity is preserved. The argument view wins. Well not entirely. Norton overlooks part of Bishop's original argument. It's that we misunderstand our history of science unless we take both Einstein's and Bohr's thought experiments to be tokens of the same thought

113. Norton, "Why Thought Experiments do Not Transcend Empiricism," 49.
114. Bokulich, "Rethinking Thought Experiments," 285; Häggqvist, "A Model for Thought Experiments," 61; Cooper, "Thought Experiments," 332; Swirski, *Of Literature and Knowledge: Explorations in Narrative Thought Experiments, Evolution and Game Theory*, 101.
115. Bishop, "Why Thought Experiments Are Not Arguments," 535–538.
116. Bishop, 538-541.
117. Norton, "Why Thought Experiments do Not Transcend Empiricism," 63.

experiment type.[118] That is, the physics community accepted Bohr's triumph over Einstein. But they did that only if Bohr showed that Einstein botched the thought experiment. And Bohr showed that only if he replicated the thought experiment, but did it correctly. So, since the physics community did accept Bohr's triumph, Bohr replicated the thought experiment; that is, he produced a second token of that same thought experiment type. Now, to rectify the oversight, Norton might re-describe. Consider two possibilities. First, Bohr replicated nothing. Rather, he carried out a similar thought experiment. Its outcome differed. The physics community thought it better. Bohr triumphed. There are two thought experiments. They differ. And each is a different argument type. Alternately, second, Bohr did replicate Einstein's thought experiment. Both were the same thought experiment type. They're also the very same argument type. For the conclusion is disjunctive. The physics community thought Bohr's disjunct the better one. He triumphed. Now, how could Bishop in turn reply? Gerrymandering! That is, to the second possibility, that's not the intuitive notion of an argument, and, so, it smells of one specially devised to get around his objection. To the first, when giving the history, we call it "the Clock in a Box thought experiment," not "Einstein's Clock in a Box thought experiment" or "Bohr's Clock in a Box thought experiments," and so, again, your re-description isn't natural and, consequently, smells of being devised only to get around the objection. Finally, this possible reply lacks justification insofar as, to so insist on a natural or intuitive form of description over another one hardly itself escapes a charge similar to that of gerrymandering.

Now, to make my point, consider another line of response. Again, for Norton, they, thought experiments, justify beliefs however arguments do—because that's what they are, arguments. We can understand this answer in a new way. To give you a glimpse of this new way, consider again Michael Bishop's objection, which goes something like this. Sometimes we carry out a thought experiment one time and get one result—but then do it again and get another. They're repeatable. Arguments, however, aren't like this; identical arguments must have the same conclusion. So it's false that thought experiments are arguments, and so Norton's answer isn't any good. Now, both Bishop and Norton understand this answer as a thesis like $Water = H_2O$. My idea is to understand it is as description, namely, that we model them not only on experiments but arguments as well. The argument then is that, once we recognize this other model, in certain cases, there's nothing to explain. Or, better, there's no more to explain than how we can use arguments to gain knowledge about the world. If we understand matters this way, Bishop's objection misfires, since a thought experiment, like the clock in a box one, need not be repeatable. It need not because this possibility comes from modelling them on experiments, and it isn't provided for when we model on arguments. What is provided for, by contrast, is the possibility that thought experiments can be distinguished by their conclusions, a possibility we actualize when so distinguishing named ones. For instance, Galileo's falling bodies thought experiment has *this* conclusion, which you cannot deny but may disagree with, if you model it on arguments. If on experiments, by contrast, you may draw a contradictory one. Even Bishop's leading example works this way. If on arguments, Bohr's clock in a box thought experiment was better than Einstein's, and the scientific community recognized it; if on experiments, Bohr repeated the clock in a box thought experiment, and the community recognized the conclusion he drew to be correct, unlike that which Einstein drew. Again, here, a gerrymandering objection arises, insofar as we use, naturally, an experiment model when describing the historical case. And, again, this objection lacks justification insofar as, to insist on a natural form of description over another one hardly itself escapes a charge similar to that of gerrymandering.

Finally, here is the main point. In light of this line of response, we can see, in the original debate, two models at work, the experiment one and the argument one, and, in particular, that the experiment one needn't

118. Bishop, "Why Thought Experiments Are Not Arguments," 540.

be fundamental. To require that we always model thought experiments on experiments and then infer that they have no outcomes essentially or that, since they're repeatable, their outcomes are accidental, is to unjustly subjugate the use of another model, the argument one in particular.

In this light, we can more clearly see that Elgin's arguments unjustly favour a way of thinking about thought experiments and that an objection, based upon them, that thought experiment outcomes are necessarily accidental, does not, as it stands, succeed.

3.4.2 An Overall Difference: Complexity

From this description of an outcome difference, let us turn to that of an overall difference. It's that stories in works of literary fiction tend to be more complex than thought experiments. This obvious difference isn't trivial. My main task here will be to explain why. In short, it's that, unlike the stories, being surveyable enters into what thought experiments are.

The plan is, first, to summarize Sorensen's argument that, roughly, the complexity of novels precludes their being thought experiments. Then I'll raise objections and, in light of them, recast Sorensen's argument in terms of surveyability—thereby explaining why the difference isn't trivial.

Sorensen's Complexity Argument

Here is a reconstruction of Sorensen's argument. As we saw in §2.1.1, experiments are procedures for answering or raising a question about a relationship between variables by varying some of these variables and seeing what if anything happens to the others.[119] Thought experiments are experiments, which are not executed but which nevertheless purport to achieve their aims.[120] That is, a thought experiment is a kind of procedure, one presented as answering or raising a question about the relationship between certain variables, not by varying some of them and tracking any response that may occur in the others, but by thinking about varying such variables and so on. Furthermore, thought experiments must be complex but not too complex or, alternately, sufficiently detailed but not too detailed.[121] They must be complex because procedures are complex acts.[122] But they must not be too complex because "the details should support, rather than engulf, the experimental intention."[123] Now, why are we more reluctant to describe a story as a thought experiment the longer it is? Because the longer it is the more theoretically irrelevant detail it has and the less likely it is to satisfy a necessary condition on being a thought experiment. The condition is that *x* aim principally to answer a theoretical question by manipulating certain variables. So stories, such as *Crime and Punishment*, aren't thought experiments, although they can be "read as" such, e.g., as pitting "Christianity against utilitarianism."[124]

Three Internal Criticisms

First, Sorensen overlooks that his definition allows raising a question, not just answering one. *Crime and Punishment* raises many. This calls for justification, since the restriction raises the standard for what counts as relevant detail. Second, even if the definition were to require aiming to answer a question, it still wouldn't require, as he does in this argument, that the answer be a theoretical point. *Crime and Punishment*—arguably,

119. Sorensen, *Thought Experiments*, 186.
120. Sorensen, 205.
121. Sorensen, 224.
122. Sorensen, 211.
123. Sorensen, 224.
124. Sorensen, 223.

see below—makes many non-theoretical ones. This again calls for justification, since it too raises the standard. Third, even if the definition were to require aiming at a theoretical answer, it still wouldn't require that this be its principal or dominant aim. *Crime and Punishment*—again, arguably—has many non-dominant aims. This too raises the standard and so calls out for justification.

Two External Criticisms

First, *Crime and Punishment*, the flagship example, on some scholarly interpretations, aims primarily to answer a theoretical question. For example, literary scholar Joseph Frank argues, in effect, that all Dostoevsky's late novels do so. Consider:

> [In *Notes from Underground*] Dostoevsky has also at last found the great theme of his later novels, which will all be inspired by the same ambition to counter the moral-spiritual authority of the ideology of the radical Russian intelligentsia (depending on whatever nuance of that ideology was prominent at the time of writing). In this respect, the nucleus of Dostoevsky's novels may be compared to that of an eighteenth-century contes philosophiques, whose characters were also largely embodiments of ideas; but…they will be fleshed out with all the verisimilitude and psychological density of the nineteenth-century novel of Social Realism and all the dramatic tension of the urban-Gothic roman-feuilleton.[125]

On *Crime and Punishment* in particular:

> The aim of [the radical intelligentsia's] ideas, as [Dostoevsky] knew, was altruistic and humanitarian, inspired by pity and compassion for human suffering. But these aims were to be achieved by suppressing entirely the spontaneous outflow of such feelings, relying on reason… to master all the contradictory and irrational potentialities of the human personality, and… encouraging the growth of a proto-Nietzsche egoism among an elite of superior individuals to whom the hopes of the future were to be entrusted. Raskolnikov… was created to exemplify all potentially dangerous hazards contained in such an ideal, and the moral-psychological traits of his character incorporate this antinomy between instinctive kindness, sympathy, and pity on the one hand, and on the other, a proud and idealistic egoism that has become perverted into a contemptuous disdain for the submissive herd. [126]

In short, Raskolnikov's ideal leads, against feeling, to the crime, among other evils, and the primary aim of the novel is to exemplify what is potentially dangerous in this ideal, i.e., in these theories of the then radical intelligentsia.

Second, the complexity isn't irrelevant detail but elaboration. We saw the idea in Davies' conclusion, in §3.2.1. Noël Carroll wields it against Sorensen:

> They [i.e., artworks like E.M. Forster's novel *Howards End*] are, in a word, more concrete than routine philosophical thought experiments, and this concreteness, in turn, is connected to their effectiveness in stimulating ethical understanding [i.e., sharpening ethical concepts by, he argues, means of enthymemes and an array of character contrasts]. 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.[127]

That is, certain artwork's extra detail, because concrete, helps, on his account, to achieve its theoretical aim, i.e., ethical understanding; thus, since the extra detail isn't irrelevant, Sorensen's argument fails.

Elgin argues similarly. First, she argues that, on her account, nothing prevents our construing the works as elaborate thought experiments. That is, since there's no reason to deny that an extended and, specifically, elaborate, as opposed to an austere, thought experiment, fairly free from any particular theory, can afford "epistemic access" to certain normally inaccessible aspects of the world, i.e., normative, psychological and

125. Chapter 30, Frank, *Dostoevsky: A Writer in his Time*.
126. Chapter 34, Frank.
127. Carroll, "The Wheel of Virtue: Art, Literature, and Moral Knowledge," 18.

metaphysical ones, we can construe works of literary fiction as elaborate thought experiments and, specifically, as affording this same access.[128] Second, she argues that the detail of works can elaborate thought experiments. That is, from Kathleen Wilkes' idea that we don't know what to think in response to certain austere thought experiments,[129] Elgin argues that we do know in response to certain works that elaborate them. As she puts it:

> Metaphysical thought experiments are often science fictional. Some are so austere that in their philosophical settings we do not know what to think. Literary and cinematic fictions help us out. What should we make of Putnam's brains in a vat? The Matrix supplies an answer. What would a computer that passed the Turing test be like? His name is Hal.[130]

Third, and finally, she replies to an objection—very roughly Sorensen's argument—that since "stereotypical" thought experiments tend to be "austere" and, as noted above, univocal, we've "a reason to deny that works of fiction are thought experiments."[131] She suggests that we infer, instead, that "some thought experiments are more austere than others," since (i), as argued, "there is a continuum of cases from Maxwell's demon and trolley problems through the myth of the cave and *Emile* to 'didactic fictions' like *Animal Farm* and *Uncle Tom's Cabin*, to *Middlemarch* and *Oedipus Rex*" and (ii) it's doubtful that strictly speaking we can sharply distinguish thought experiments from works of literary fiction.[132]

By the way, Elgin also brushes off the objection. That is, she replies that "demarcating the boundary is not so important" and, accordingly, that, even if we cannot "call works of fiction thought experiments," we can nevertheless construe the ones as the others and, in particular, her main claim still stands, i.e., "that fictions, thought experiments, and standard experiments function in much the same way."[133] This matters for the response I'll now give.

Recasting Sorensen's Argument

To introduce the main point around which I'll recast Sorensen's argument, the difference in complexity isn't merely one in the quantity of detail. Rather, as I said in §3.2.1, it's also in how hard it is to take in, recall and describe—i.e., in non-surveyability.[134] The main point is that to say, "these stories in works of literary fiction are more complex than those thought experiments," or the like, means that we cannot easily survey the stories.

My argument for this claim is that, as a matter of course, unlike "story in a work of literary fiction" and the like, we use the expression "thought experiment" and its cognates as we use "proverb" or "anecdote." That is, normally, being easy to take in, to recall and to repeat governs our use of the one as it does the others. Take, for example, a paradigm proverb like "a bird in hand…" It moves easily from person to person, and we expect others to as well. That is, we're surprised at obscure ones, frustrated when unable to recall one that's "on the tip of my tongue," assume them to be a common currency, and so on. We decline or hesitate to call a "proverb" that which doesn't play these roles in our lives. Similarly, paradigm thought experiments like Newton's Bucket and Thomson's Violinist pass easily from person to person, and again we expect others to as well. That is, we expect to get the gist of the whole in a short time, not in weeks or months and only after second or third readings, and to recall the whole and, if asked, to explain it fairly well without reference or much delay, without getting flummoxed by details, and so on. We decline or hesitate to call a "thought experiment" that which doesn't play these roles.

128. Elgin, "Fiction as Thought Experiment," 232.
129. Wilkes, *Real people: Personal Identity Without Thought Experiments*.
130. Elgin, "Fiction as Thought Experiment," 236.
131. Elgin, 239.
132. Elgin, 240.
133. Elgin, 240.
134. Cf. Wittgenstein on "A mathematical proof must be perspicuous" (Part III, §§1–2, Wittgenstein, *Remarks on the Foundations of Mathematics*, 143).

A subsidiary point is that we may, after briefly hesitating, justify calling it so by appeal to how it is nevertheless like an experiment, and, more to the point, if we do not notice we're switching models in doing so, we may easily overlook all of this, i.e., that being surveyable has anything to do with what thought experiments are. If we do notice, by contrast, we might qualify that we're regarding or construing it as a thought experiment.

In light of these points, I'll recast Sorensen's argument as follows. We often call something a thought experiment, as a matter of course, only if it is not complex and, specifically, only if it is surveyable, i.e., easy to take in, to recall, and to repeat. This replaces his necessary condition, that X is a thought experiment only if it aims principally to answer a theoretical question by manipulating certain variables; and, this aim is the principle one only if X isn't complex, i.e., only if there isn't too much irrelevant detail. Next, stories in works of literary fiction tend to be complex, that is, non-surveyable. This replaces his claim that novels, such as *Crime and Punishment*, have so much detail irrelevant to any theoretical aim that this aim cannot be the principle one. Finally, whereas we tend to recognize thought experiments in virtue of their being simple—i.e., surveyable—this is not the case for the stories; and, to regard the stories as thought experiments is to risk losing our grip on them as normally appreciated, i.e., to turn to an exceptional case or introduce a new one. This replaces Sorensen's conclusion that, although we can read such novels as thought experiments, they are not literally so.

What of the internal objections? None apply insofar as I neither rely on his definition nor his necessary condition. What of the two external ones? The Frank interpretation of Dostoevsky's late novels, on which they have a primary theoretical aim, is no counterexample. So the first doesn't apply. The second doesn't either. That is, complexity may not be irrelevant detail, because it's elaboration, but even elaboration reduces surveyability. In this light, Davies undermines his own conclusion tentatively calling some stories, such as *1984*, "elaborate thought experiments," instead of, say, suggesting a way to explain away the difference in complexity. Also, in this light, Carroll, letting extra detail in the door, risks disqualifying artworks, like *Howards End*, as thought experiments, even if it's concrete and so relevant to and helpful for the ethical task. Finally, Elgin, making that first move, overlooks that elaboration of a thought experiment may dissolve it, and so there is a reason to think that elaboration may preclude its affording such epistemic access. Her third move, moreover, overlooks that seeing a mere difference in austerity, and with it singularity of outcome, between thought experiments, as well as seeing a continuum between them and works of literary fiction, papers over this important difference in surveyability.

What of Elgin's reply that demarcating isn't so important and that, even if we can't call the works thought experiments, her account goes through? First, on the importance of demarcating, Davies objected that Elgin doesn't respond well to Lamarque and Olsen's anti-cognitivist challenge, touched on above, insofar as she doesn't explain how the learning she describes counts as being properly literary. As he puts it, "it is not clear why such testing is properly viewed as integral to our engagement with literary fictions as literature, and thus why we are entitled to view literature as a source of knowledge rather than as a source of hypotheses."[135] Elgin responded, in short, that "such policing of disciplinary boundaries strikes me as ill-advised."[136] So far as I understand the response, it misses the point. I want to understand how I learn from stories in works of literary fiction as I ordinarily read them, and it is not at all clear whether or not I'm getting an answer unless it's clear that the account concerns such stories so read. The demarcation, whether sharp or not, then, matters insofar as I want to know I'm getting such an answer instead of talking about something else. Again, the worry is that, if

135. Davies, "Learning Through Fictional Narratives in Art and Science," 63.
136. Elgin, "Fiction as Thought Experiment," 239.

we construe the stories as thought experiments—i.e., see them as functioning the same way—we lose our grip on the stories and it's not longer clear we're getting the answer we want.

These remarks bring this chapter and my overall project to its end. The project, summarized in my preface, brings a Wittgensteinian-inspired approach to bear upon questions about the nature of thought experiments and how we learn from stories in works of literary fiction. If successful, I've made some progress toward clearing away philosophical problems arising from those questions.

Bibliography

Appreciate v. 3a. In *OED Online*. 2019.

Atwood, Margaret. "Afterward." Chap. 48 in *The Handmaid's Tale*. Audible & Harcourt, 2017.

Austen, Jane. *Pride and Prejudice*. IBooks ed. Edited by James Kinsley. Oxford: Oxford UP, 1813.

Baker, Alan. "Quantatative Parsimony and Explanatory Power." *British Journal for the Philosophy of Science* 54 (2003): 245–259.

Baker, G.P., and P.M.S. Hacker. *Wittgenstein, Understanding and Meaning: Volume 1 of an Analytic Commentary on the Philosophical Investigations, Part II*. 2nd ed. (extensively revised by P.M.S. Hacker). Oxford, UK: Blackwell, 2005.

Balint, Benjamin. *Kafka's Last Trial: The Case of a Literary Legacy*. New York: W. W. Norton & Company, 2018.

Benjamin, Walter. *The Work of Art in the Age of Mechanical Reproduction*. Edited by Hannah Arendt. Translated by Harry Zohn. Schocken/Random House, 1936.

Berenstain, Jan, and Stan Berenstain. *The Berenstain Bears and the Trouble with Friends*. Reprint. Random House Books for Young Readers, 1987.

Berkeley, George. *Principles of Human Knowledge and Three Dialogues Between Hylas and Philonous*. Edited by Roger Woolhouse. London: Penguin, 2004.

Bishop, Michael. "Why Thought Experiments Are Not Arguments." *Philosophy of science* 66, no. 4 (1999): 534–541.

Black, Max. "The Identity of Indiscernibles." *Mind*, 1952, 153–164.

Bokulich, Alisa. "Rethinking Thought Experiments." *Perspective on science* 9, no. 3 (2001): 285–307.

Borges, Jorge Luis. *Collected Fictions: Jorge Luis Borges*. Translated by Andrew Hurley. New York: Penguin, 1999.

Brown, James R. *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. 2nd ed. New York: Routledge, 2011.

———. "Thought Experiments: A Platonic Account." In *Thought Experiments in Science and Philosophy*, 119–128. Savage, Md.: Rowman & Littlefield, 1991.

———. "Why Thought Experiments Trascend Empiricism." In *Contemporary Debates in the Philosophy of Science*, edited by Christopher Hitchcock, 23–43. Malden, MA: Blackwell, 2004.

Brown, James R., and Yiftach Fehige. "Thought Experiments." *Stanford Encyclopedia of Philosophy*, 2019. https://plato.stanford.edu/entries/thought-experiment.

Carroll, Noël. "Introduction." In *Theories of Art Today*, edited by Noël Carroll, 3–24. The University of Wisconsin Press, 2000.

———. "The Wheel of Virtue: Art, Literature, and Moral Knowledge." *Journal of Aesthetics and Art Criticism*, 2002, 3–26.

Conrad, Joseph. *The Nigger of the 'Narcissus'*. Heinemann, 1897.

Cooper, Rachel. "Thought Experiments." *Metaphilosophy* 36, no. 3 (2005): 328–347.

Currie, Gregory. *The Nature of Fiction*. Cambridge: Cambridge University Press, 1990.

Daniel, Nolan. "Quantitative Parsimony." *British Journal for the Philosophy of Science for the philosophy of science* 48 (1997): 329–343.

Davenport, Edward A. "Literature as Thought Experiment (On Aiding and Abetting the Muse)." *Phil. Soc. Sci.* 13 (1983): 279–306.

Davies, David. "Fiction." In *Routledge Companion to Aesthetics*, edited by B. Gaut & D. Lopes, 263–73. London: Routledge, 2000.

———. "Learning Through Fictional Narratives in Art and Science." In *Beyond Mimesis and Convention*, edited by Roman Frigg and Matthew Hunter, 51–69. Springer, 2010.

———. "Thought Experiments and Fictional Narratives." *Croatian Journal of Philosophy* 19 (2007): 29–45.

Diamond, Cora. *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*. Cambridge, Mass.: MIT Press, 1995.

———. "What If X Isn't the Number of Sheep? Wittgenstein and Thought-Experiments in Ethics." *Philosophical Papers* 31, no. 3 (2002): 227–250.

Donner, Wendy. *The Liberal Self: John Stuart Mill's Moral and Political Philosophy*. 66–82. New York: Cornell UP, 1991.

Dostoevsky, Fyodor. *Demons: A Novel in Three Parts*. Vintage Classics ed. Translated by Richard Pevear and Larissa Volokhonsky. New York: Kindle, 1995.

———. *The Brothers Karamazov: A Novel in Four Parts with Epilogue*. Translated by Richard Pevear and Larissa Volokhonsky. New York: Farrar, Straus / Giroux, 1990.

Duhem, Pierre. *La Théorie Physique: Son Objet, Sa Structure*. Chevalier et Rivière, 1906.

Einstein, Albert. *Relativity: The Special and the General Theory*. Translated by Robert W. Lawson. New York: Random House, 1961.

Einstein, Albert, and Léopold Infeld. "The Evolution of Physics." *Revue de Métaphysique et de Morale* 1:46 (1939): 173–173.

Elborough, Travis, and Helen Gordon. *Being a Writer: Advice, Musings, Essays and Experiences from the World's Greatest Authors*. London: Quarto Publishing, 2017.

Elgin, Catherine. "Fiction as Thought Experiment." *Perspective on Science* 22, no. 2 (2014): 221–241.

———. "The Laboratory of the Mind." In *A Sense of the World: Essays on Fiction, Narrative, and Knowledge*, edited by Wolfgang Huemer John Gibson and Luca Pocci, 43–54. Oxford: Routledge, 2007.

Eliot, George. *Middlemarch*. Edited by Rosemary Ashton. London: Penguin Classics, 1994.

Forster, Michael. "Wittgenstein on Family Resemblance Concepts." In *Wittgenstein's Philosophical Investigation: A Critical Guide*, edited by Arif Ahmed. Cambridge University Press, 2010.

Frank, Joseph. *Dostoevsky: A Writer in his Time*. Edited by Mary Petrusewicz. New Jersey: Princeton UP, 2012. Kindle.

Gaddis, William. *A Frolic of his Own*. IBooks ed. New York: Simon & Schuster, 1994.

———. *Agapē Agape*. New York: Penguin, 2002.

———. *Carpenter's Gothic*. Harmondsworth: Penguin, 1985.

———. *JR*. Champaign: Dalkey Archive Press, 1971.

———. *The Recognitions*. Champaign: Dalkey Archive Press, 1952.

Gale, Richard. "On Some Pernicious Thought-Experiments." In *Thought Experiments in Science and Philosophy*, edited by Tamara Horowitz and Gerald Massey, 297–304. Savage, Md.: Rowman & Littlefield, 1991.

Galilei, Galileo. *Two New Sciences: Including Centers of Gravity and Force of Percussion*. 2nd ed. Edited and translated by Stillman Drake. Toronto: Wall & Emerson, 1989.

Gaut, Berys. "'Art' As a Cluster Concept." In *Theories of Art Today*, edited by Nöel Carroll, 25–44. University of Wisconsin Press, 2000.

Gendler, Tamar Szabó. "Galileo and the Indispensability of Scientific Thought Experiment." *British Journal for the Philosophy of Science* 49 (1998): 397–424.

———. *Thought Experiment: On the Powers and Limits of Imaginary Cases*. New York: Garland Press, 2000.

———. "Thought Experiments Rethought—and Reperceived." *Philosophy of Science* 71, no. 5 (2004): 1152–1163.

———. "Tools of the Trade: Thought Experiments Examined." *The Harvard Review of Philosophy* Spring (1994).

Gibbs, Laura. *Aesop's Fables*. Oxford University Press, 2002.

Gibson, John. *Fiction and the Weave of Life*. Oxford: Oxford University Press, 2007.

Glock, Hans-Johann. "Wittgenstein on Concepts." In *Wittgenstein's Philosophical Investigation: A Critical Guide*, edited by Arif Ahmed, 88–108. Cambridge University Press, 2010.

Hagberg, Garry L. "Metaphor." In *The routledge companion to aesthetics*, 2nd ed., edited by Berys Gaut & Dominic McIver Lopes, 371–382. London: Routledge, 2005.

Häggqvist, Sören. "A Model for Thought Experiments." *Canadian Journal of Philosophy* 39, no. 1 (2009): 55–76.

Horowitz, Tamara, and Gerald Massey. "Introduction." In *Thought Experiments in Science and Philosophy*, edited by Tamara Horowitz and Gerald Massey, 1–30. Rowman & Littlefield, 1991.

Ichikawa, Jonathan, and Benjamin Jarvis. "Thought-Experiment Intuitions and Truth in Fiction." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 142, no. 2 (2009): 221–246.

Interpret, v. 1. a. In *OED Online*. 2019.

Interpret, v. 1. b. In *OED Online*. 2019.

Irvine, Andrew. "Thought Experiments in Scientific Reasoning." In *Thought Experiments in Science and Philosophy*, edited by Tamara Horowitz and Gerald Massey, 149–166. Lanham: Rowman & Littlefield, 1991.

Jackson, Frank. "Epiphenomenal Qualia." *Philosophical Quarterly* 32 (1982): 127–136.

Joyce, James. *Ulysses*. Edited by Jeri Johnson. Oxford World's Classics. Oxford UP, 1922.

Kafka, Franz. *The Metamorphosis*. Translated by Stanley Corngold. New York: Bantam Classic, 2004.

Kerouac, Jack. *On the Road*. New York: Viking Press, 1957.

Kivy, Peter. *Philosophies of Arts: An Essay in Differences*. Cambridge University Press, 1997.

Klagge, James C. "The Wittgenstein Lectures." In *Ludwig Wittgenstein: Public and Private Occasions*, edited by J. Klagge and A. Nordmann, 331–372. Oxford: Rowman & Littlefield, 2003.

———. "Wittgenstein and von Wright on Goodness." *Philosophical Investigations* 31, no. 3 (2018): 291–303.

Kripke, Saul. *Naming and Necessity*. Cambridge, Mass.: Harvard UP, 1972.

Kuhn, Thomas. "A Function for Thought Experiments." In *The Essential Tension: Selected Studies in Scientific Tradition and Change*, 240–265. The University of Chicago Press, 1977.

Lamarque, Peter. "Literature." In *The Routledge Companion to Aesthetics*, 2nd ed., edited by Berys Gaut and Dominic McIver Lopes, 571–584. Routledge, 2005.

Lamarque, Peter, and Stein Haugom Olsen. *Truth, Fiction, and Literature: A Philosophical Perspective*. Clarendon, 1996.

Laymon, Ronald. "Thought Experiments of Stevin Mach and Gouy: Thought Experiments as Ideal Limits and as Semantic Domains." In *Thought Experiments in Science and Philosophy*, edited by Tamara Horowitz and Gerald Massey, 167–192. Lanham: Rowman & Littlefield, 1991.

Lewis, David. *Counterfactuals*. Cambridge, Mass.: Harvard UP, 1973.

Mach, Ernst. "On Thought Experiments." In *Knowledge and Error: Sketches on the Psychology of Enquiry*, translated by Paul Foulkes, 134–147. Reidel Publishing Company, 1976.

Max, D.T. "Every Ghost Story is a Love Story: A Life of David Foster Wallace," 2012. Kindle.

McComb, Geordie. "Thought Experiment, Definition, and Literary Fiction." In *Thought Experiments in Philosophy, Science, and the Arts*, edited by Mélanie Frappier, Letitia Meynell, and James Robert Brown, 207–222. New York: Routledge, 2013.

McDowell, John Henry. "Are Meaning, Understanding, etc., Definite States?" In *The Engaged Intellect: Philosophical Essays*, 79–95. Cambridge, Mass.: Harvard UP, 2009.

———. "How Not to Read *Philosophical Investigation*: Brandom's Wittgenstein." In *The Engaged Intellect: Philosophical Essays*, 80–96. Cambridge, Mass.: Harvard UP, 2009.

Melville, Herman. *Moby Dick or the Whale*. Pennsylvania: The Franklin Library, 1851.

Meynell, Letitia. "Imagination and Insight: A New Account of the Content of Thought Experiments." *Synthese* 191 (2014): 4149–4168.

Mill, John Stuart. *Utilitarianism*. Peterborough, Ontario: Broadview Press, 2000.

Miščević, Nenad. "Mental Models and Thought Experiments." *International Studies in the Philosophy of Science* 6, no. 3 (1992): 215–226.

———. "Modelling Intuitions and Thought Experiments." *Croatian Journal of Philosophy*, no. 20 (2007): 181–214.

Mohanty, J. N. "Method of Imaginative Variation in Phenomenology." In *Thought Experiments in Science and Philosophy*, edited by Tamara Horowitz and Gerald Massey, 261–272. Lanham: Rowman & Littlefield, 1991.

El-Mohtar, Amal. "March's Book Club Pick: 'The Power', by Naomi Alderman." *New York Times Online*, Oct. 25, 2017.

Munch, Robert. *Marilou Cass-Cou*. Éditions Scholastic, 2001.

Nagel, Thomas. "What Is It Like To Be a Bat?" *The Philosophical Review* 83:4 (1974): 435–450.

Nersessian, Nancy. "In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling." In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 291–301. 1992.

———. "Thought Experimenting as Mental Modeling: Empiricism Without Logic." *Croatian Journal of Philosophy* 20 (2007): 125–161.

New, Christopher. "Walton on Imagination, Belief and Fiction." *The British Journal of Aesthetics* 36 (1996): 159–165.

Norton, John. "Thought Experiments in Einstein's Work." In *Thought Experiments in Science and Philosophy*, edited by Tamara Horowitz and Gerald J Massey, 129–144. Savage, Md.: Rowman & Littlefield, 1991.

———. "Why Thought Experiments do Not Transcend Empiricism." In *Contemporary Debates in Philosophy of Science*, edited by Christopher Hitchcock, 44–66. Blackwell, 2004.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.

Nussbaum, Martha. "Finely Aware." In *Love's Knowledge: Essays on Philosophy and Literature*. Oxford: Oxford UP, 1990.

Orwell, George. *Animal Farm & 1984*. Orlando: Harcourt, 2003.

Orwell, George. "Politics and the English Language." In *The Broadview Anthology of Expository Prose*. Toronto: Broadview Press, 2011.

Plato. "Plato: Complete Works," edited by John M. Cooper and D. S. Hutchinson. Indianapolis/Cambridge: Hackett, 1997.

Proust, Marcel. *À la Recherche du Temps Perdu*. Intégral "Les 7 Tomes" ed. 1922. Kindle.

Rawls, John. *A Theory of Justice*. Cambridge, Mass.: Harvard UP, 1971.

Reid, Thomas. *An Inquiry into the Human Mind on the Principles of Common Sense*. Edited by Derek R. Brookes. University Park, Pennsylvania: Pennsylvania State UP, 1997.

Rescher, Nicholas. "Thought Experimentation in Presocratic Philosophy." In *Thought Experiments in Science and Philososphy*, 31–42. Savage, Md.: Rowman & Littlefield, 1991.

Searle, John. "Proper Names." *Mind* 67 (1958): 166–173.

Sorensen, Roy A. *Thought Experiments*. New York: Oxford UP, 1992.

Strawson, P.F. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen, 1959.

Styron, William. *Sophie's Choice*. New York: Vintage International, 1976.

Suits, Bernard. *The Grasshopper: Games, Life and Utopia*. Peterborough, Ont.: Broadview Press, 2005.

Swirski, Peter. *Of Literature and Knowledge: Explorations in Narrative Thought Experiments, Evolution and Game Theory*. London and New York: Routledge, 2007.

Thomson, Judith Jarvis. "A Defense of Abortion." *Philosophy & Public Affairs*, 1971, 47–66.

Thought Experiment. In *OED Online*. 2019.

Tolstoy, Leo. *Anna Karenina*. Translated by Richard Pevear and Larissa Volonkhonsky. New York: Penguin Classics, 2000.

———. *War and Peace*. Edited by Amy Mandelker. Translated by Louise and Aylmer Maude. Oxford: Oxford UP, 2010.

Virgil. *The Aeneid of Virgil: In the Verse Translations of John Dryden*. Translated by John Dryden. Pennsylvania: The Franklin Library, 1975.

Wallace, David Foster. *Infinite Jest*. 10th Anniversary ed. New York: Back Bay Books, 2006.

———. *Infinite Jest*. 20th Anniversary ed. New York: Back Bay Books, 2016. Kindle.

———. "The Salon Interview: David Foster Wallace." In *Conversations with David Foster Wallace*, edited by Stephen J. Burn. Interview by Laura Miller. Jackson: University Press of Mississippi, 2012. Kindle.

———. *This is Water: Some Thoughts, Delivered on a Significant Occasion, About Living a Compassionate Life*. New York: Little Brown & Company, 2009.

———. "Track Three." In *David Foster Wallace: In His Own Words*. Hachette Audio, 2014.

Walton, Kendall L. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, Mass.: Harvard University Press, 1990.

Wharton, Edith. *Age of Innocence*. Oxford: Oxford UP, 2006.

Wilkes, Kathleen V. *Real people: Personal Identity Without Thought Experiments*. Oxford: Clarendon, 2003.

Williamson, Timothy. *The Philosophy of Philosophy*. Malden, MA: John Wiley & Sons, 2008.

Wittgenstein, Ludwig. *Culture and Value*. Translated by Peter Winch. University of Chicago Press, 1984.

———. *Philosophische Untersuchungen: Philosophical Investigations*. 3rd ed. Translated by G.E.M. Anscombe. Malden, MA: Blackwell, 2001.

———. *Philosophische Untersuchungen: Philosophical Investigations*. 4th ed. Translated by G.E.M Anscombe, P.M.S. Hacker, and Joachim Schulte. Chichester, West Sussex: Wiley-Blackwell, 2009.

———. *Remarks on the Foundations of Mathematics*. 3rd ed. Edited by G. E. M and Anscombe, Rush Rhees, and G. H. von Wright. Oxford: Blackwell, 1978.

———. *The Blue and Brown Books*. Malden, MA: Blackwell, 1958.

———. *Tractatus Logico Philosophicus*. Edited by C.K. Ogden Trans. Project Gutenberg, 2010.

Wright, G.H. Von. *Varieties of Goodness*. London: Routledge, 1963.

Zola, Émile. *Le Roman Expérimental*. 5th ed. Edited by G. Charpentier. Paris: G. Charpentier, 1881. PDF.