THE JOURNAL OF AESTHETICS AND ART CRITICISM

DAVID EGAN

Literature and Thought Experiments

ABSTRACT

Like works of literature, thought experiments present fictional narratives that prompt reflection in their readers. Because of these and other similarities, a number of philosophers have argued for a strong analogy between works of literary fiction and thought experiments, some going so far as to say that works of literary fiction are a species of thought experiment. These arguments are often used in defending a cognitivist position with regard to literature: thought experiments produce knowledge, so works of literary fiction can too. This article concedes that works of literary fiction can be put to use in thought experiments, but not in a way that is helpful to the cognitivist. In particular, it draws three disanalogies in the ways we engage critically with thought experiments and with literary fictions. First, we use thought experiments to make arguments; second, we read thought experiments in strongly allegorical terms; and third, the terms of criticism we apply to thought experiments and to works of literature differ. Although these disanalogies present problems for the cognitivist position, they also give us a sharper picture of the distinctive educative potential of works of literary fiction.

Thought experiments present short fictional narratives that have well-established, if controversial, applications in physics and philosophy as well as less prominently in other fields of inquiry. The narratives in thought experiments share a number of similarities with other kinds of fictional narrative; indeed some philosophers have argued that literary fictions simply *are* thought experiments, albeit of a particular kind.[1] This argument is often paired with a defense of cognitivism in the arts. In answer to the question of how works of fiction can be sources of knowledge, defenders of the thought experiment analogy point to the epistemic value of thought experiments: thought experiments also present purely fictional scenarios and yet they can provide us with knowledge. Of course, the matter is not quite so simple: the epistemic status of thought experiments is also a matter of considerable controversy, and, as David Davies (2007, 2010) notes, moves made in the debate about the epistemic status of thought experiments find analogues in the debate about the epistemic status of literary fiction. Nevertheless, the well-established role of thought experiments in uncontroversially cognitive enterprises lends intuitive plausibility to the defense of literary fictions as having epistemic value, provided we can show a robust analogy between literary fictions and thought experiments.

This article casts doubt on the possibility of drawing a strong analogy between literary fictions and thought experiments. I begin by getting a bit clearer on what I mean by 'thought experiment' and 'literary fiction' and what sorts of claims we might want to make about the analogy between the two. Then I present three related differences in how we engage critically with thought experiments and with literary fictions. The main upshot of these disanalogies is that thought experiments have a delimited application in the service of particular arguments whereas the uses of literary fictions are less clearly defined and more diverse. I conclude with some reflections about the prospects for cognitivism if the disanalogies I identify carry weight.

I. PRELIMINARY DEFINITIONS AND DISTINCTIONS

We begin to see the intuitive plausibility of treating literary fictions as a kind of thought

experiment when we consider some broad similarities between the two. First of all, both of them are *fictional*: unlike reports of laboratory experiments, the events described in thought experiments and literary fictions have not actually taken place.[2] Second, both literary fictions and thought experiments are essentially narrative: some counterfactual reasoning simply involves imagining a nonrealized state of affairs, but narrative fictions present the imagination with a series of temporally distinct events.

We use fictional narratives for a number of quite different reasons: we construct fictional narratives when engaging in deliberative counterfactual reasoning, for instance, or when expressing wishes or fantasies.[3] We also have a practice of recounting and consuming stories for the aesthetic pleasure it brings. These are the sorts of fictional narratives I mean by 'literary fiction.' I use 'aesthetic pleasure' here in a broad sense, including everything from the simple fun of a well-spun yarn to the hard-won intellectual, spiritual, or emotional rewards that intense engagement with a dense text can produce.

Another—arguably overlapping—use of fictional narratives is in the construction of thought experiments. Thought experiments are similar in many respects to laboratory experiments: both seek to raise, make salient, or answer a question or set of questions, and they do so by exploring the relationship between two or more variables.[4] Experiments require carefully controlled conditions so that confounding factors are eliminated or consigned to the background in such a way that the relationship between the relevant variables stands out as clearly as possible. The main difference between thought experiments and laboratory experiments is that laboratory experiments are actually performed, whereas thought experiments are performed only in what James Robert Brown (2011) calls "the laboratory of the mind."

As useful examples, I will return to two famous thought experiments throughout this article. The first is Galileo's thought experiment, in which he invites us to imagine a falling cannonball attached to a musket ball. The idea is this: if heavier objects fall faster than lighter objects, the cannonball attached to the musket ball must fall both faster and slower than an unattached cannonball. Faster, because the compound cannonball–musket ball weighs more than the unattached cannonball, and slower, because the lighter, and hence slower falling, musket ball will act as a drag on the cannonball to which it is attached. Although the actual upshot of this thought experiment is controversial,[5] it is often taken to show a contradiction in Aristotelian physics: Galileo examines the relationship between weight and rate of fall and allegedly shows that, in this instance, Aristotelian physics predicts two contradictory results. The thought experiment can reveal this contradiction without any empirical investigation.

The second thought experiment is Thomson's violinist, whose narrative I reproduce here in full:

> You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you, "Look, we're sorry the Society of Music Lovers did this to you—we would never have permitted it if we had known. But still, they did it, and the violinist now is plugged into you. To unplug 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." (Thomson 1971, 48–49)

Thomson uses this narrative to draw a distinction between a purported right to life and a purported right to what is needed to sustain life. Even if the violinist has a right to life, Thomson argues, we are morally permitted to unplug ourselves and let the violinist die. It would be nice of us to keep the violinist alive, but we are not morally obligated to do so. Likewise, she reasons, a woman is morally permitted to terminate the life of an unborn fetus, even if that fetus has as much of a right to life as the violinist.

Like a good experimenter, Thomson's thought experiment carefully constrains the variables in her narrative in order to bring out with maximal clarity the relation between the variables she wishes to consider. The unconscious victim is a famous violinist so that we confer as much value as possible on this person: if we think it acceptable to detach ourselves from a cultural treasure, then it is acceptable to detach ourselves from *anybody*. This *anybody* is a stranger, not someone to

whom we already have an emotional connection. *Only you* can save the violinist: Thomson rules out by stipulation the possibility that anyone else can offer the violinist the same kidney dialysis. *You*, however, did not consent to having the violinist connected to you. Thomson also rules by stipulation that only by being uninterruptedly attached to the violinist are you in a position to save his life. In making these tight stipulations, Thomson's thought experiment is not as restrictive as some. Orphans are a staple feature in many ethical thought experiments so as to remove the extraneous element of family whose emotional ties might interfere with our moral calculations. The idea is to rule out as many extraneous variables as possible so as to focus our moral attention on the variables the thought experimenter wants to consider.

Advocates of the thought experiment analogy claim that the difference between a narrative like Thomson's and the sorts of fictional narratives we typically describe as "literary" is a difference of degree rather than kind. In particular, both thought experiments and literary fictions are used to generate reflections that—so the argument goes—generate knowledge. And indeed, we do see a striking number of similarities between the kinds of work they do. Thought experiments carefully constrain their variables so as to bring out with maximal clarity the relationship between the variables under consideration, and the same could be said for literary fictions. It is not by happenstance or whim that Cervantes pairs Don Quixote with Sancho Panza: by placing these two characters side by side, Cervantes brings out the distinctive qualities of each all the more clearly. Thought experiments and literary fictions also both make use of exemplification, where elements in their narratives serve as samples or examples of broader types, and they calibrate their narratives so that the samples they choose serve as particularly telling instances of the phenomena under investigation.[6]

What should we make of these similarities? Here are three possibilities. The first I will call the *applicability claim*: the sorts of narratives that we typically treat as literary fictions can be used in thought experiments. Second, the *cognitivism claim*: the analogies between literary fictions and thought experiments are strong enough that we can apply arguments about the cognitive value of thought experiments in a defense of the cognitive value of literary fiction. And third, the *identity claim*: literary fictions *are* thought experiments, even if not all thought experiments are literary fictions.

I think the applicability claim is true, but only in an obvious and trivial sense: philosophers *do* sometimes use works of literary fiction in their thought experiments.[7] And indeed, there seems to be nothing in the nature of literary fiction to forbid this appropriation. If thought experiments are fictional narratives that are carefully calibrated to raise, make salient, or answer a particular question or set of questions, there is no reason why a thought experimenter might not take advantage of the rich stock of narrative that world literature provides us with if one of those narratives can serve the desired purpose.

However, the truth of the applicability claim is not only obvious and trivial, it is also insufficient for the stronger cognitivism claim. A robust cognitivism about literary fiction has to show not just that literature can be a source of knowledge; it has to show that the distinctively *literary* features of literature generate knowledge. Jane Austen's novels are a source of knowledge about the fashions and mores of Regency England, but an anticognitivist would be unimpressed for at least two reasons. First, novels are not the best source of this kind of historical knowledge: historians provide historical knowledge within a framework of rigorous scholarship that makes their accounts more thorough and more reliable. And second, this sort of knowledge is incidental to the distinctively aesthetic qualities that make Austen's novels works of literature. Austen likely could not write aesthetically rich novels set in Regency England without conveying information about the fashions and the mores of the time, but then *Hamlet* cannot be performed without a stage to perform it on: the stage or the historical setting might be necessary conditions for certain aesthetic pleasures, but they are not themselves the sources of those pleasures. Literary fictions are also sometimes used as sources for bibliomancy—and only works deemed to be of high aesthetic value are so used, and used on account of their aesthetic value—but we likewise would rightly see their potential for bibliomancy as incidental to their aesthetic value.

So to get from the uninteresting applicability claim to the interesting cognitivism claim, we need to show not only that literary fictions can be used in thought experiments, but also that this

thought-experimental use of literary fictions essentially involves the literary features of the fiction. Getting from the cognitivism claim to the identity claim requires further argument to the effect that appreciating the literary features of fiction simply *is* to treat the fiction as a kind of thought experiment. Defenders of the thought experiment analogy want to defend the cognitivism claim and often the identity claim rather than simply the applicability claim, and I think both of these stronger claims are mistaken. In the next three sections I explore three important disanalogies between thought experiments and literary fictions before returning more explicitly in the final section to why I think these disanalogies undermine the stronger claims I have considered here.

The disanalogies I discuss draw on the fact that both literary fictions and thought experiments invite critical engagement and that this critical engagement brings out clearly the distinctive features of both uses of fictional narrative. Literary criticism and interpretation aims (among other things) to tease out as precisely as possible the ways in which a given fictional narrative produces its distinctive aesthetic pleasures, so considering the methods of critical engagement with literature should give us a sharper view of the distinctively literary features of literary fiction. Thought experiments also find their place within an argumentative apparatus, and considering the role they play in arguments will sharpen our view of their distinctive features. In considering disanalogies in these kinds of critical engagement, I show that literary and thought-experimental uses of fictional narratives are different, incompatible, and sometimes even at cross-purposes.

II. FIRST DISANALOGY: MAKING ARGUMENTS

We use thought experiments to make or contribute to an argument. Thought experiments serve as engines for generating particular questions or answers to particular questions that have a role in an argument. They can play a variety of different roles in an argument.[8] Galileo's thought experiment with the cannonball and musket ball allegedly helps us see a contradiction in Aristotelian physics. Thomson's violinist thought experiment seeks to undermine a rival view not by showing a contradiction, but by making clear a distinction that shows that the argument she targets is invalid. Putnam's Twin Earth thought experiment not only challenges a rival view but also lends powerful positive support to his preferred alternative. Singer's drowning child thought experiment lends support not to a theoretical position but to a practical one: if we accept his analogy and the argument it supports, it should significantly alter our behavior rather than just our theoretical commitments, assuming we do not already make significant contributions to aid organizations. As these examples suggest, thought experiments can play a diversity of roles in arguments, but in every case they play *some* role in an argument.[9]

Two further points about the relationship between thought experiments and arguments bear emphasis. First, thought-experimental narratives are *essentially* involved in arguments. A fictional narrative only becomes a thought experiment to the extent that it is deployed in an argument; otherwise it is just an intriguing narrative. Thomson's narrative occupies less than a page in a longer article, in which she speaks in propria persona and spells out explicitly how her thought experiment pertains to her argument. Qua thought experiment, her narrative is essentially tied to this further discussion, since this discussion spells out the narrative's experimental purpose. This point emerges more clearly if we consider the parallel case with empirical experiments. Newton's famous experiments with prisms are experiments only because he used their results in developing a theory of color. Without this broader argumentative purpose, his "experiment" would have simply been a charming light show. Both thought experiments and their real-world equivalents become experiments only insofar as they are roped into a larger argumentative structure. And it is only in their contribution to this larger argumentative structure that they get their distinctive cognitive payoff.

Second, the purpose of a thought experiment is *exhausted* in making or contributing to an argument. Whatever other aesthetic qualities that narrative might contain are irrelevant to the purposes of the thought experimenter. At best, they make the thought experiment more vivid or more compelling, but this vividness and compellingness is useful to the thought experimenter only insofar as it contributes fruitfully to the argument. Like Wittgenstein's proverbial ladder, we can, as it were, throw away our thought experimental narrative once its work is done. This is not to say that thought experiments cannot be repurposed

and reused—generating variations on the trolley problem has become a minor industry within the philosophical profession—but all of these repurposings and reuses remain in service of the arguments into which they are incorporated.

If thought experiments are exhaustible in this way, their role in arguments is also fungible. Thomson could have substituted another thought experiment for her violinist thought experiment provided it drew the same distinction just as compellingly. Nothing in her argument requires this particular story about the violinist. It happens to serve her argumentative purpose very well, but its value rests in how well it serves this purpose. Another thought experiment—or an argument of any kind—that does the same work could be substituted in its place.

The aesthetic pleasures of literary fiction are not entirely divorced from argumentation—I discuss this point further in Section IV, below—but literary fictions have a different relation to argument than thought experiments. When we read literary fictions, we do not have such a clear and delimited purpose in mind as making a particular move in a particular argument. For one thing, no single purpose could adequately account for our diverse reasons for reading literature, let alone account for them essentially, as argument does for thought experiments. Furthermore, none of these purposes exhausts the use of a literary fiction or make it fungible.[10] I may derive amusement from a story by P. G. Wodehouse, but I do not regard the story simply as an engine for generating amusement and think that any other equally amusing story could have done just as well. Even with a purpose as simple as pure amusement, the uniqueness of just *this* story remains.

Most importantly, we do not settle the purpose of reading a work of literature in advance of the reading itself. I might have some idea of what to expect the first time I open *The Brothers Karamazov*, but to the extent that I read with a mind to addressing only a delimited set of questions framed in a particular way, I have cripplingly constrained the aesthetic pleasure I can derive from it. We read literature with a degree of openness: we remain alert and attentive to the ways the text might surprise us, provoke unexpected thoughts, insights, feelings, and so on. None of this is required, or even expected, when we read thought experiments. Reading a thought experiment might provoke thoughts, insights, or feelings beyond what is needed for the argument at hand, but these additional responses are accidental outcomes rather than central to the experience of reading a thought experiment.

Another way of putting the same point is that, provided we remember how a thought experiment runs, we have no reason to reread it. If we read a thought experiment solely with a mind to its contribution to an argument, the only reason to reread it is to refresh our memory of how the argument runs or to engage with it critically. By contrast, rereading is a common feature of literary engagement with a text. Rich literary texts are especially multifarious, and every new reading can garner fresh insights. We reread thought experiments in order to refresh ourselves on the *same* argument, but we reread literary fictions with the expectation of discovering something *new*.

As I noted in the previous section, I accept the applicability claim: literary fictions can be used in thought experiments. For instance, philosophers discussing death often use *The Death of Ivan Ilyich* as a thought experiment.[11] Like Thomson's violinist thought experiment, Tolstoy's novella can be used to make salient a distinction we might otherwise overlook, in this case between two conceptions of knowing that one is going to die. In one sense, Ivan Ilyich has long known perfectly well that he is going to die, but he is nevertheless surprised when he learns he is going to die: it seems two different conceptions of "knowing one is going to die" are at work here. However, this way of reading *The Death of Ivan Ilyich*—reading Tolstoy's novella as a thought experiment that makes salient an important distinction—is different from the more open-ended experience of reading *The Death of Ivan Ilyich* without any delimited purpose in mind. When we use it as a thought experiment in this way, Tolstoy's novella exhausts its purpose once it has made a particular distinction salient, and we could just as well have used some other thought experiment provided it made the same distinction equally salient. But when we approach *The Death of Ivan Ilyich* as a work of literary fiction, it is neither so readily exhausted nor so readily fungible.

I do not mean to condemn the use of *The Death of Ivan Ilyich*, or other works of literature, as thought experiments. If this is the distinction you want to draw, by all means use Tolstoy to draw it. This thought-experimental use of Tolstoy may offer only a limited engagement with the text, but

as long as one does not claim that this more limited reading is *all* Tolstoy's novella can offer us, no aesthetic foul is committed by using it in this way. *War and Peace* is a great novel, but it can also be a great paperweight.

III. SECOND DISANALOGY: ALLEGORICAL READING

The second disanalogy I draw between reading narratives as thought experiments and reading them as literary fictions is that our reading of thought experiments treats them allegorically, or more particularly as what Northrop Frye (1957, 90–91) calls naive allegory. We read the concrete elements of a thought experiment as having a one-to-one correspondence with a particular set of abstract concepts. The concrete story in the text is like a map whose purpose is exhausted by helping us see our way more clearly through the abstract territory it maps. Reading literary fiction can certainly spur abstract thinking as well, but the concrete elements of the narrative remain irreducibly a part of our imaginative engagement.

Thought experiments are concerned with abstract or general relations rather than with particulars. Galileo is not interested in the relation between cannonballs and musket balls but in the relation between mass and acceleration due to gravity; Thomson is not interested in the relation between famous violinists and kidney ailments but in the relation between certain rights and obligations. The concrete elements in a thought experiment make an arguably ineliminable contribution to exploring these relations: they make salient the relations under examination with a clarity that would arguably be impossible in a purely abstract treatment. However, even if they are ineliminable in this way, these concrete elements are at best a means to an end, namely, the end of thinking clearly about relations that hold at a more general or abstract level.

In this respect, we read thought experiments allegorically: we treat each element in the story as a symbolic representation of some aspect of the abstract problem under consideration. Not all allegorical readings make thought experiments of their objects, of course. Neoplatonists developed an influential reading of the *Odyssey* as an allegory for the soul's journey. I would hesitate to call this allegorical reading a thought experiment, though, because it does not read the *Odyssey* as part of an argument. The *Odyssey* might be read so as to *illustrate* a certain conception of the soul's journey, but this reading does not contribute to an *argument* in support of this conception. Thought-experimental readings, then, are naïve allegories—allegories whose every concrete element has an allegorical analogue at the abstract level—that contribute to an argument.

By contrast, outside the strictest allegorical readings, the concrete particularities of narratives are irreducible parts of what we attend to when we read a narrative as literature. In contrast with naive allegory, Frye notes that even such a "continuous allegory" as *The Faerie Queen* "is still a structure of images, not of disguised ideas, and commentary has to proceed with it exactly as it does with all other literature, trying to see what precepts and examples are suggested by the imagery as a whole" (Frye 1957, 80). In other words, literary reading—the sort of reading that seeks to maximize aesthetic pleasure—draws meaning from the connections *between* elements at the concrete level rather than finding meaning only at the allegorical level. These concrete particularities, then, cannot be straightforwardly reduced to abstract ideas.

The concrete particularities of a literary work are irreducible for two opposite reasons. First, the concrete particularities of a thought experiment point univocally to a particular set of abstract or general reflections, whereas the abstract and general reflections a literary engagement with a text might prompt have no defined limit. And second, although our literary engagement with a text *might* prompt a wide range of abstract and general reflections, it does not *have to* prompt any such reflections at all. I will discuss each of these points in turn.

Literary interpretation obviously can generate more general and abstract reflections in the form of symbolism, metaphor, and so on. Interpreters of Kafka's *The Metamorphosis* have read Gregor Samsa's transformation as a metaphor for late-capitalist ennui, father/son psychological trauma, humanity's uncomfortable relation with its own animality—or indeed as a turn-of-the-century *Sprachkrise*-influenced attack on the very possibility of metaphorical meaning altogether. A good work of literary fiction is unboundedly generative in this way. And, as a result, it is not allegorical in the way that a thought experiment is. Because the interpretive possibilities are not fixed

in advance, we cannot abstract away from the concrete particularities of Kafka's narrative if we want to think about it as a work of literary fiction.

Furthermore, literary reading does not require abstract or general reflections. I can derive great interest and pleasure from reading *Pride and Prejudice* as a story about Elizabeth Bennett, Mr. Darcy, and others without drawing any more general views about the nature of class or marriage. By contrast, if I do not infer from Thomson's narrative about the violinist to abstract reflections about the right to life then I have failed to read it as a thought experiment. Reading a thought experiment aims at something beyond what we aim at when reading for pleasure.

This treatment of thought experiments as devaluing the concrete particulars of a narrative in favor of a narrow range of abstract reflections might seem to undervalue the imaginative possibilities of engaging in thought experiments. In arguing that at least some works of literature can be treated as thought-experimental "virtue wheels," Noël Carroll (2002) argues that a careful reading of a novel is required to extract the appropriate understanding of the relation between the virtues treated in the novel. One might object to Carroll's argument here by saying that he stretches the definition of a thought experiment too far, subsuming under the category of thought experiments interpretive procedures that make no specifiable contribution to a particular argument. However, I think a different response to Carroll is more apt: in trying to read literary texts as thought experiments, he treats them allegorically, as a thought experimenter should, thereby limiting our interpretive response. In Carroll's reading of *Howards End*, the characters in the novel instantiate the virtues of imagination and practicality in varying degrees, which is helpful because it prompts "the audience to apply concepts of virtue and vice to the characters, thus exercising and sharpening their ability to recognize instances of these otherwise vaguely defined or highly abstract concepts" (Carroll 2002, 13). Like a thought experimenter, Carroll treats the concrete particularities of the novel as means to the end of abstract reflection. To the extent that we might want to fill out in our imagination the world that E. M. Forster presents to us, we do so only to get sharper on the abstract concepts its characters instantiate. These characters, then, are of interest primarily as instantiations of bundles of abstract concepts. Carroll treats the world of the novel not as an autonomous source of reflection, which could lead us in any one of a number of different directions, but as a mine from which to extract and smelt a delimited set of abstract concepts.

In the previous section, I insisted that there was nothing *wrong* with using a work of literature to generate answers to particular questions. I could likewise argue here that there is nothing wrong with treating *Howards End* as a thought experiment for generating reflections on the relation between the virtues of imagination and practicality, provided of course that we do not conclude that this is *all* that *Howards End* is good for. Carroll is welcome to his allegorical reading of *Howards End*, but the novel retains other interpretive possibilities.[12]

However, with regard to the question of how carefully we engage with the concrete particularities of a narrative, a thought-experimental reading might not always be so innocent. We can prompt ethical reflection by reading a narrative either as a thought experiment or as a work of literary fiction, and the choice of reading shapes the kind of reflection we pursue. From a variety of positions, philosophers have argued that moral reasoning requires a detailed attention to particulars that cannot be codified in abstract formulations. This kind of thinking leads Jonathan Dancy (1985) and Marilyn Friedman (1987) to question the applicability of thought experiments in moral philosophy and leads Martha Nussbaum (1990) and Cora Diamond (1991) to claim that literature is a source of moral insight that cannot be reduced to philosophical argument.[13] This is not the place to take a stand on this matter, but I note that the abstraction of a thought-experimental reading is arguably detrimental rather than beneficial to generating moral insight from literary narratives.

IV. THIRD DISANALOGY: TERMS OF CRITICISM

The third disanalogy I want to draw between our engagement with thought experiments and with literary fictions is that we use a different vocabulary when engaging with them critically. Although there is some overlap in terms of criticism, the differences are pronounced and important. Where our critical engagement with literary works is primarily aesthetic, our critical engagement with thought experiments is more in line with how we

engage critically with arguments. The difference in the forms of critical response strongly suggests that we engage in two different kinds of reading when reading thought experiments and when reading literature.

To get clearer on the ways in which one might respond critically to a thought experiment, let me return to Thomson's violinist. The argument in "A Defense of Abortion" has generated a tremendous amount of discussion and debate, a great deal of it focused on the thought experiment about waking up plugged into a famous comatose violinist. Although most of her critics agree with Thomson that the subject of the story is morally permitted to unplug herself from the violinist, Singer (2011, 132–134) disputes this point on utilitarian grounds. Other critics dispute the analogy between the violinist and the unborn fetus. For instance, outside of rape, a woman engages voluntarily in the sexual act that results—however unintentionally—in pregnancy, whereas the violinist is hooked up to the subject of the story without the subject's knowledge or tacit consent. The woman's voluntary involvement gives her a special responsibility toward the fetus that the subject of Thomson's story does not have. Or Thomson's analogy fails because the violinist is a stranger, whereas the intimate relation a woman has to a fetus in her body confers special obligations. Or again, abortion involves killing a fetus whereas unplugging the violinist only involves letting the violinist die, and there is a morally salient difference between killing and letting die.[14]

All of these objections—and the attendant responses by Thomson and her allies—concern not Thomson's story but the way that she applies this story in the interpretive frame that constitutes the bulk of her paper. If this criticism applies to Thomson's story, it applies to the extent that her story is a bad analogy to the case of abortion: if she wanted to make her argument stick, her critics would conclude, she should have composed a different story. All of this criticism presupposes the instrumentality of the thought experiment: it is an engine for generating a salient distinction, and its success or failure hinges on how it performs the task that Thomson stipulates it is meant to perform.

In these respects, the kinds of criticism we apply to thought experiments are very similar to the kinds of criticism we apply to arguments. Thought experiments are components of arguments, and we might challenge the validity of the argument or challenge the author's claim that the thought experiment makes the contribution to the argument that she claims it does. Just as we might question the premises of an argument, we might challenge the coherence of the narrative setup of a thought experiment or, like Singer, dispute what the author of the thought experiment takes it to show. In addition, thought experiments are subject to their own distinctive forms of criticism. One prominent form of criticism that is unique to thought experiments is the counter thought experiment, where the critic poses a variant on the original thought experiment, which casts doubt on the validity of the original thought experiment (see Brown 2007).

By contrast, we almost never see thought experiments criticized on aesthetic grounds. None of Thomson's critics allege that the character of the hospital director in her narrative lacks clear motivation or is clichéd or two-dimensional. No one is inclined to spill much ink over praising or criticizing the elegance or compactness or precision of Thomson's prose. Furthermore, Thomson or her allies could respond to such criticism in a way that makes clear the different dimensions of criticism involved in criticizing a thought experiment and a work of literary fiction. Fine, Thomson could reply, I am no Tolstoy, but nothing you have said has so much as put a dent in my argument.

Broadly speaking, we have two different categories of criticism at work here: argumentative and aesthetic. The former concerns itself with notions like validity, truth, and justification, and the latter concerns itself with a much broader palette, including notions of style, nuance, and psychological acuity. However, these two categories of criticism might seem to overlap more than I have suggested. I will address one spurious case of seeming overlap before considering more substantial claims to overlap.

The spurious case of overlap concerns terms of criticism that we do in fact see in both argumentative and aesthetic criticism, but which are applied in very different ways. A critic might attack either a thought experiment or a work of literary fiction on grounds of implausibility, for instance. But if we look at this criticism more closely, we see a marked difference in its tenor. If we complain that the narrative setup of a thought experiment is implausible, we are complaining that this implausibility compromises the validity or the clarity of the argument. By contrast, calling a work of literary fiction implausible is generally part of a

complaint that the narrative is contrived or violates the narrative logic of the work in such a way as to diminish our overall enjoyment of or investment in the story.

That said, under certain circumstances, argumentative criticism can find justifiable application as part of a literary engagement with a text. With regard to argumentative criticism in literary engagement, imagine Stalin reading *Animal Farm*. He might suggest that the analogy to Stalinism is false or that Orwell paints a misleading picture of Soviet communism. Granted, the text of *Animal Farm* says nothing about Stalinism explicitly, but Orwell's target is clear enough. We might be over-reading if we take *Animal Farm* to be offering anything resembling a deductive argument against Stalinism, but the analogy itself is clear enough, and the criticism of Stalinism is hardly veiled. Nor, I think, does reading a criticism of Stalinism into *Animal Farm* divert us from the distinctive literary features of the novel. On the contrary, any responsible reading of *Animal Farm* should have something to say about the analogy to Stalinism. More generally, literary texts can be polemical or argumentative in all sorts of ways, and we bracket their argumentative upshot at our peril.

One familiar anticognitivist response to this point is that the argumentative *upshot* of a work of literary fiction is different from an *argument*. We may be able to extract an argument against Stalinism from *Animal Farm*, but that argument is not stated explicitly, nor is it defended with the sort of argumentative apparatus that we would expect of a serious, rigorous argument. The argumentative work of *Animal Farm* is mostly left to the reader: if we extract an argument from a work of literary fiction and then proceed to criticize it, we are criticizing something other than the work of literary fiction itself. Our argumentative criticism of *Animal Farm* would at best target claims we have come to entertain because we read *Animal Farm*, not *Animal Farm* itself.

This anticognitivist argument is not without its critics,[15] but it is also stronger than what I need for my more modest aim: I want to show only that the analogy with thought experiments does not hold. And although it may be true that a certain kind of argumentative criticism is not only appropriate to literary fictions, but that it engages precisely with the literary features of the fictions, I do not think this is true of the particular kind of argumentative criticism that treats the work of fiction as a thought experiment.

To make this point, I consider an argument to the contrary. In arguing that many literary fictions serve as thought experiments that give us conceptual clarity in thinking about virtue, Carroll (2002, 17–18) claims that aesthetic and emotional responses to a text are essential sources of this clarity. But Carroll's reading of *Howards End* makes thought-experimental virtues of literary vices and implicitly does the reverse. On Carroll's reading, it is a narrative strength to have each character in *Howards End* clearly instantiate a particular combination of the virtues of imagination and practicality. By contrast, a critic with more literary inclinations might complain about just this sort of clear instantiation: these characters are not *characters*, such a critic might allege, but two-dimensional stand-ins for certain philosophical ideas. Carroll writes approvingly that "the novel parades before us a series of characters who instantiate these virtues in varying degrees" (Carroll 2002, 12), but we could imagine another critic using exactly these words in criticism of the novel.

More generally, the features that make for a good thought experiment often make for a bad story and vice versa. Where we might praise a thought experiment for clearly schematizing the abstract concepts it aims to treat, we might criticize a work of literature for being too schematic. Where we might praise a work of literature for its subtlety and nuance, we might criticize a thought experiment for being muddy and lacking precision. We look for singularity of focus and clear analogies between the concrete and the abstract in thought experiments, but in literary fictions we want a broader perspective and one not so readily reduced to a set of abstract relations. The features of a fictional narrative that make it laudable as a work of literature tend to get in the way if we want to treat it as a thought experiment. Some forms of argumentative criticism can have a place in literary criticism, but not the sort of argumentative criticism we apply to thought experiments.

Even a polemical novel like *Animal Farm* is far less schematic than a thought experiment, and this is because its aims and effects are much broader. Thought experiments raise, make salient, and/or answer a particular question or delimited set of questions. Their purpose is exhausted by this extranarrative result or set of results. By contrast, we cannot exhaustively list the question or questions

that *Animal Farm* raises. Although the polemical intent of the novel is clear in its broad strokes, there is no single, definite "moral" to the story that could be stated in propositional terms. Interpreting Orwell's polemic is not a matter of unspooling it into the form of an argument. Grasping even just the polemical thrust of the novel requires irreducibly literary terms of criticism, even if some terms of argumentative criticism are also apposite.

V. CONCLUDING THOUGHTS ON COGNITIVISM

In the first section of this article, I outlined three claims we might want to draw from the apparent analogy between literary fictions and thought experiments. I happily granted the first (the applicability claim) as trivially true: philosophers and others are welcome to make use of literary fictions in constructing thought experiments. But I cast doubt on two stronger claims: the cognitivism claim—we can defend the cognitive value of literary fictions by drawing analogously on the cognitive value of thought experiments—and the identity claim—literary fictions *are* thought experiments. The last of these claims entails the other two: if the identity claim is true, then so is the cognitivism claim (and, clearly, so is the applicability claim); and hence if the cognitivism claim is false, then so is the identity claim. For that reason, I will focus on the ways in which my disanalogies make problems for the cognitivism claim.

I claimed that the cognitivism claim is a stronger claim than the applicability claim because the cognitivism claim requires not just that a work of literary fiction *can* be a source of knowledge, but that it must be a source of knowledge by virtue of its distinctively literary features. The disanalogies I have drawn have focused on the ways in which we engage critically with literary fictions and thought experiments because it is presumably in this critical engagement that we can best hope to draw whatever epistemic benefit literary fictions and thought experiments provide. The knowledge-producing features of thought experiments (such as they are—as I have noted, the question of whether and how thought experiments produce knowledge is also controversial) pertain to their argumentative role in leading to general or abstract truths. But my argument has targeted precisely these points of analogy between literary fictions and thought experiments. In treating a fictional narrative as a thought experiment, we treat it as a part of a particular argument whose purpose is exhausted by the role it plays in that argument; we treat it as a "naive allegory," where each element in the narrative stands in relation to a particular abstract concept; and we apply to it critical tools that hold schematic clarity as a virtue. In all of these respects, the critical engagement we bring to thought experiments not only differs from the critical engagement we bring to literary fictions, but is at cross-purposes to the aim of finding aesthetic pleasure in literary fiction.

If the cognitivism claim fails for these reasons, then so does the identity claim. This point bears emphasis because the thought experiment analogy is not wedded to a defense of cognitivism. It would be interesting if the identity claim were true regardless of its consequences for cognitivism. And conversely, my argument hardly means lights out for the cognitivist. The analogy with thought experiments is just one strategy in the cognitivist's playbook, and others might well be successful. My primary aim in this article has not been to attack cognitivism, but rather to attack an analogy—one that is often deployed in a defense of cognitivism—that I think is unhelpful.

This argument should not be seen as limiting our sense of what literature can do. On the contrary, the disanalogies I have drawn repeatedly aim to show that literature can do much *more* than what thought experiments do. One feature of treating a work of literature as a thought experiment is that it instrumentalizes it: it sees the real value of the work of literature in the epistemic achievements it helps us toward. To the extent that a work of literature is a means to some other end, it is also fungible: it could in principle be replaced by any other set of epistemic tools that do the same work. Without making a strong argument for the intrinsic value of literature, we can nevertheless acknowledge that works of literary fiction are not quite so readily instrumentalized as thought experiments. Thought experiments have a limited and clearly defined use. Literary fiction, to the extent that it can be instrumentalized, at least has a much broader range of possible uses.

This broader range could well include other educative possibilities. In particular, various so-called neo-cognitivist approaches suggest ways in which we might learn from literature and other arts besides instilling truth and knowledge.[16] Without giving us anything as direct as propositional

knowledge, works of literature can train us in various epistemic virtues—a careful attention to particulars, flexibility in our thinking, a deeper appreciation of the significance of the knowledge we already have or the practices we engage in, and so on—that have a distinct kind of epistemic value.

Of course, we might want to resist instrumentalizing works of literary fiction even this far: training us in epistemic virtues is not what literature is *for* any more than literature is a source of thought experiments to be used in arguments. However, it seems this less sharply defined set of epistemic virtues is less susceptible to the danger I noted with regard to thought experiments, of pushing to one side the distinctively *literary* features of works of literary fiction.[17]

DAVID EGAN
Society of Fellows
University of Chicago
Chicago, Illinois, 60637

INTERNET: degan@uchicago.edu

REFERENCES

Benn, S. I. 1992. "Wickedness." In *Ethics and Personality: Essays on Moral Psychology*, edited by John Deigh, 191–206. University of Chicago Press.

Brown, James Robert. 2007. "Counter Thought Experiments." *Royal Institute of Philosophy Supplement* 61: 155–177.

———. 2011. *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, 2nd edition. New York: Routledge.

Brown, James Robert, and Yiftach Fehige. 2014. "Thought Experiments." *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/thought-experiment/.

Carroll, Noël. 2002. "The Wheel of Virtue: Art, Literature, and Moral Knowledge." *The Journal of Aesthetics and Art Criticism* 60: 3–26.

Conolly, Oliver, and Bashshar Haydar. 2007. "Literature, Knowledge, and Value." *Philosophy and Literature* 31: 111–124.

Dancy, Jonathan. 1985. "The Role of Imaginary Cases in Ethics." *Pacific Philosophical Quarterly* 66: 141–153.

Davenport, Edward A. 1983. "Literature as Thought Experiment (on Aiding and Abetting the Muse)." *Philosophy of the Social Sciences* 13: 279–306.

Davies, David. 2007. "Thought Experiments and Fictional Narratives." *Croatian Journal of Philosophy* 7: 29–45.

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

Diamond, Cora. 1991. "Anything but Argument?" In *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*, 291–308. MIT Press.

———. 2002. "What if X Isn't the Number of Sheep? Wittgenstein and Thought-Experiments in Ethics." *Philosophical Papers* 31: 227–250.

Elgin, Catherine Z. 2007. "The Laboratory of the Mind." In *A Sense of the World: Essays on Fiction, Narrative, and Knowledge*, edited by John Gibson, Wolfgang Huemer, and Luca Pocci, 43–54. London and New York: Routledge.

———. 2014. "Fiction as Thought Experiment." *Perspectives on Science* 22: 221–241.

Friedman, Marilyn. 1987. "Care and Context in Moral Reasoning." In *Women and Moral Theory*, edited by Eva Feder Kittay and Diana T. Meyers, 190–204. Totowa, NJ: Rowman & Littlefield.

Frye, Northrop. 1957. *Anatomy of Criticism: Four Essays*. Princeton University Press.

Gibson, John. 2008. "Cognitivism and the Arts." *Philosophy Compass* 3: 573–589.

Goodman, Nelson. 1968. *Languages of Art*. Indianapolis: Hackett.

Kagan, Shelly. 2012. *Death*. Yale University Press.

Kuhn, Thomas S. 1977. "A Function for Thought Experiments." In *The Essential Tension: Selected Studies in Scientific Tradition and Change*, 240–265. University of Chicago Press.

McMahan, Jeff. 2002. *The Ethics of Killing: Problems at the Margins of Life*. Oxford University Press.

Norton, John D. 1996. "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy* 26: 333–366.

———. 2004. "Why Thought Experiments Do Not Transcend Empiricism." In *Contemporary Debates in Philosophy of Science*, edited by Christopher Hitchcock, 44–66. Oxford: Blackwell.

Nussbaum, Martha C. 1990. *Love's Knowledge: Essays on Philosophy and Literature*. Oxford University Press.

Rowe, M. W. 1997. "Lamarque and Olsen on Literature and Truth." *Philosophical Quarterly* 47: 322–341.

Singer, Peter. 2011. *Practical Ethics*, 3rd edition. Cambridge University Press.

Sorensen, Roy A. 1992. *Thought Experiments*. Oxford University Press.

Swirski, Peter. 2007. *Of Literature and Knowledge: Explorations in Narrative Thought Experiments, Evolution, and Game Theory*. London and New York: Routledge.

Thomson, Judith Jarvis. 1971. "A Defense of Abortion." *Philosophy & Public Affairs* 1: 47–66.

---

1. Examples include Davenport (1983), Carroll (2002), Swirski (2007), and Elgin (2007, 2014).

2. Or at least, if and to the extent that these narratives are true, fidelity to their truth is not what guides the author's constructive activity. Their truth is inessential to their narrative purpose.

3. Telling lies might also count as narrative fiction—indeed, probably the most common form of narrative fiction—although it differs from these other instances in that the teller deliberately tries to conceal the fact that the narrative is fictional.

4. This rough definition borrows significantly from Sorensen's (1992, 186) definition of "experiment" and his related definition of "thought experiment" (205).

5. Kuhn (1977) shows that the story is more complicated than the simple version I tell here. I use Galileo's thought experiment in this article as a toy example, and I do not think anything in my argument hinges on the fact that what I offer oversimplifies the history.

6. Elgin (2014) discusses the similar role of exemplification in experiments, thought experiments, and literary fiction, drawing on the work of Goodman (1968). The final

section of her article also provides a helpful catalogue of the kinds of similarities that might support the thought experiment analogy—a catalogue far more extensive than the brief summary provided in this paragraph.

7. Carroll (2002, 9) develops this point, drawing in particular on Benn's (1992) invocation of Iago and Milton's Satan against the Socratic claim that no one knowingly does evil.

8. What follows is not meant to be anything like an exhaustive taxonomy. We find more thorough attempts at providing a taxonomy for thought experiments in Brown (2011) and Brown and Fehige (2014).

9. Exactly what role thought experiments play in arguments is a matter of some controversy. Norton and Brown represent far ends of a spectrum: Norton's empiricism sees thought experiments as nothing more than ordinary arguments, "disguised in a vivid pictorial or narrative form" (Norton 2004, 45; see also Norton 1996), whereas Brown's Platonism maintains that thought experiments are an autonomous source of a priori knowledge (see especially Brown 2011). My argument tries to remain as neutral as possible with regard to the question of just what relationship thought experiments have to arguments.

10. Conolly and Haydar (2007) argue for some degree of fungibility between literary fictions in their defense of a form of propositional cognitivism: if the knowledge we gain from a work of literature can be expressed in propositional form, then those propositions can be separated from the work itself. Although I am skeptical of this view, I do not need to engage it fully to make my point here. Even if the *cognitive* payoff of a story can be parceled into countably distinct propositions, this only amounts to the wholesale fungibility of a work of literature if the cognitive payoff of a story exhausts its aesthetic payoff. And that seems highly doubtful.

11. See, for example, Kagan (2012, 193–194).

12. It is unclear how far Carroll himself is committed to the claim that works of literary fiction simply *are* thought experiments—how far, that is, he advocates the identity claim I outlined in Section I, as distinct from the cognitivism claim. At a number of points, he writes about "literary thought experiments" as if the phrase were synonymous with "works of literary fiction" (for example, "literary thought experiments such as *Howards End* rely on the audience to note the relevant conceptual discriminations and to reach the pertinent conclusions"; Carroll 2002, 14). At other points, he seems more reticent (for example, Carroll 2002, 23n74).

13. Diamond (2002) offers a qualified case for how thought experiments can be useful in moral reasoning while also highlighting the potential for misuse.

14. McMahan (2002, 362–398) discusses each of these three objections in turn.

15. See, for example, Rowe (1997).

16. Gibson (2008, 585–586) provides an overview of neo-cognitivist approaches.

17. The article emerged from a paper on Wittgenstein's use of language-games that I presented at the University of Chicago's Literature and Philosophy Workshop in the spring of 2015. I am grateful to the participants for their feedback and for helping me to see that this argument about thought experiments warranted a paper of its own. I am also grateful to the feedback I received from the referees for this journal and from Cathal Ó Madagáin.