

Explicitism about Truth in Fiction

William D'Alessandro

Abstract

The problem of truth in fiction concerns how to tell whether a given proposition is true in a given fiction. Thus far, the nearly universal consensus has been that some propositions are 'implicitly true' in some fictions: such propositions aren't expressed by any explicit statements in the relevant work (or works), but are nevertheless held to be true in those works on the basis of some other set of criteria. I call this family of views 'implicitism'. I argue that implicitism faces serious problems, whereas the opposite view ('explicitism') is much more plausible than has previously been thought. After mounting a limited defense of explicitism, I explore a difficult problem for the view and discuss some possible responses.

1. Truth in Fiction

An old problem in the philosophy of fiction is how to tell which propositions are true in, or true according to, a given fiction. A solution to this problem is a theory of 'truth in fiction'.

Questions about truth in fiction impinge on major issues in ontology (e.g., the existence and properties of fictional characters), metaphysics (e.g., fictionalist approaches to domains of discourse such as modality and mathematics), and philosophy of language (e.g., the semantics and logic of claims about 'truth according to a source'), among other subjects. So the problem is one that many philosophers ought to care about.

This paper attempts to shed some light on the problem of truth in fiction, in two ways. First, I distinguish between two kinds of theories of truth in fiction, which I'll call *explicitism* and *implicitism*. According to explicitism, the set of propositions that are true in a given fiction is

a subset of the set of propositions that are expressed by explicit statements in the corresponding works. Implicitism, meanwhile, is the denial of explicitism. According to implicitism, there are some works, and some propositions which aren't expressed by any explicit statements in those works, such that the propositions in question are nevertheless true in the corresponding fiction. (What exactly is meant by these definitions will be clarified shortly.) Although some version of the explicitism/implicitism distinction has been observed by other authors, the present paper offers some clarifications about the two views and surrounding issues that haven't been made elsewhere.

It's safe to say that implicitism is the dominant way of thinking about truth in fiction (although there's been lively debate over exactly which version of the theory is correct).¹ In fact, no philosopher that I know of has seriously questioned implicitism or attempted to defend any version of explicitism. My second and primary goal here is to challenge this orthodoxy. I hope to show that implicitism is much less obviously true, and that explicitism is much less obviously false, than the standard wisdom suggests.

¹ See, for example, David Lewis, 'Truth in Fiction', *American Philosophical Quarterly* 15 (1978), 37-46; Kendall Walton, *Mimesis as Make-Believe* (Cambridge, MA: Harvard University Press, 1990); Gregory Currie, *The Nature of Fiction* (New York: Cambridge, 1990); Alex Byrne, 'Truth in Fiction: The Story Continued', *Australasian Journal of Philosophy* 71 (1993), 24-35; Derek Matravers, 'Beliefs and Fictional Narrators', *Analysis* 55 (1995), 121-122; Christopher New, 'A Note on Truth in Fiction', *Journal of Aesthetics and Art Criticism* 55 (1997), 421-423; John F. Phillips, 'Truth and Inference in Fiction', *Philosophical Studies* 94 (1999), 273-293; Richard Hanley, 'As Good as It Gets: Lewis on Truth in Fiction', *Australasian Journal of Philosophy* 82 (2004), 112-128.

I begin by raising a problem for implicitism about truth in fiction. Roughly, the issue is that the purported existence of implicit truths conflicts with the possibility of sequels (or other appropriately related works) in which pretty much anything can turn out to be true, provided only that what's established in the later works is consistent with what's *explicitly* true in the earlier works. If the argument against implicitism is decisive, as I think it may be, then it's time to take a closer look at alternative approaches.

To this end, I consider whether explicitism is really as indefensible as many authors have claimed. I show that it isn't. On the contrary, explicitism is preferable to implicitism in important respects, and the most common and seemingly plausible attempts to refute it are unsuccessful. There remains one problem with the view, however, that's more difficult to resolve. The problem arises from much the same sort of reasoning that figures in the argument against implicitism, with much the same upshot. One possible response to the problem is to accept a nihilistic version of explicitism according to which there are no (nontrivial) truths in fictions. I leave it open whether this is the correct response, or whether some more robust form of explicitism might yet be salvaged.

Let me fix some terminology before going further. Throughout the paper I use the term 'work' to mean a literary artifact, such as a novel, short story or play. (I limit myself to literary fiction for convenience and simplicity, though I think that almost everything I say can probably be adapted to fiction in other media. The main challenge will be to work out a notion of explicitness that's appropriate for the medium in question.) If you like, a work can be identified with an ordered set of sentence-types. By contrast, I use the term 'fiction' (or sometimes 'story') to mean the (nonactual) state of affairs described by some work or family of works. A fiction,

then, can be identified with a set of propositions. According to this picture, to say that a proposition p is true in a fiction F is to make the set-membership claim $p \in F$.

Here's a bit more terminology. An 'explicit statement' in a work W is a declarative sentence-type appearing somewhere in W . An 'explicit proposition' in W is a proposition expressed by an explicit statement in W .² Finally, when I speak of the set of explicit propositions associated with a given work, I really mean the closure of this set under logical consequence, and probably under the consequence relation of the relevant natural language too. (Thus, properly but more clumsily speaking, explicitism is the view that all the truths in a given fiction are either expressed by explicit statements in the relevant works, or else are implied by such propositions.)

2. Implicitism about Truth in Fiction

'Does Sherlock Holmes have a third nostril?' (That is, 'Is the proposition expressed by "Holmes has a third nostril" true in the Holmes stories?') This is the sort of question to which implicitism and explicitism might give contrary answers, since the claim that Holmes lacks a third nostril is arguably true in the stories though it's nowhere made explicitly. For example, David Lewis writes:

² This is a slight abuse of language, since propositions aren't 'in' works, in the sense of set membership. But there seems to be no way to express this important idea that's both suitably compact and non-misleading, and so I've chosen to err on the side of compactness. I doubt that any serious confusion will result.

I claim it's true, though not explicit, in the [Sherlock Holmes] stories that Holmes doesn't have a third nostril; that he never had a case in which the murderer turned out to be a purple gnome; that he solved his cases without the aid of divine revelation; that he never visited the moons of Saturn; and that he wears underpants.³

The main dispute among implicitist philosophers like Lewis is about what conditions, exactly, are necessary and sufficient for a (non-explicit) proposition to be true in a work of fiction. Some implicitists are relatively liberal about these conditions, others relatively conservative. Lewis's account, for example, is on the liberal side, as it allows for many non-explicit 'background' truths to be supplied by the beliefs of the author's community. Other authors have proposed other criteria for identifying implicit truths in fiction: see, for example, the works by Walton, Currie, Byrne, Matravers, New, and Phillips cited above. I lack the space to discuss the details of any of these proposals here, unfortunately. But the argument I'll soon give against implicitism applies equally to any version of the view.

Before going further, it's worth asking what sorts of considerations have been offered in favor of implicitism. David Lewis, whom one might justly call the founding father of the view in its modern form, seems in fact to offer no argument at all. He simply asserts that there are implicit truths in fiction, such as that Holmes wore underpants and lacked a third nostril. Lewis apparently takes these claims to be so obviously true as to require no comment or justification.

Lewis's successors have often been less dogmatic, although they haven't been as thorough as they might be. At least, however, some recent authors have distinguished implicitism

³ Lewis, 'Truth in Fiction', 41.

from some form of explicitism, and they've recognized the bare possibility that explicitism might be true. For example, Phillips considers

the view that what is true in a story is what is explicitly mentioned in the text; nothing more and nothing less. I shall label this view the *Literal View*. The *Literal View*, as we have already seen, appears to be an inadequate theory. The *Literal View* fails to account for the truth of statements such as 'Iago is duplicitous', 'Tom Sawyer is a human being', 'King Lear had a consort', 'Holmes has a blood type', and countless others like them. Intuitively, there are many propositions that are part of a story which aren't expressed by any sentences in the text. A better theory, then, would be one which could account for the truth of statements of this kind.⁴

And Byrne asks:

Why not identify what is true in a fiction with what is explicitly stated in the fiction (or follows deductively from what is explicitly stated)? Well, in some fictions there are deluded narrators, and so they speak falsely. Therefore the proposal doesn't give a sufficient condition. But it doesn't give a necessary condition either. There are many truths in fiction which aren't explicitly stated, and aren't entailed by what is explicitly stated. It's true in the Holmes stories—as Lewis pointed out—that

⁴ Phillips, 'Truth and Inference in Fiction', 275.

Holmes doesn't have a third nostril, and that he never visited the moons of Saturn. However, neither of these propositions is explicitly stated in the stories, or entailed by what is explicitly stated.⁵

Although Phillips and Byrne go further than Lewis, their arguments are unsatisfying in a couple respects. For one, the version of explicitism attacked by Phillips and Byrne is a simplistic and implausible one. What Phillips calls the Literal View (and what Byrne identifies without naming) is the thesis that the set of truths in a given fiction is *coextensive* with the set of propositions expressed by explicit statements in the relevant works. As Byrne points out, this view is easily refuted, for instance by the existence of unreliable narrators. But explicitism comes in other and more credible forms. A sophisticated explicitist will hold that the set of truths in a given fiction is a *subset*—in general, a proper subset—of the set of explicit propositions in the relevant works. This view is entirely compatible with the existence of untruthful narrators, misinformed characters, and the like.

Moreover, Phillips' and Byrne's arguments (as well as Lewis's) lean heavily on unsupported appeals to implicitist intuitions. While I don't deny that many people have such intuitions, this fact alone isn't likely to move someone who has principled reasons for doubting the truth of implicitism. In the next section I offer just such a reason. (Later, in section 5, I argue that explicitism is in fact compatible with implicitist-style intuitions. So the worry here is even less pressing than it might have seemed.)

⁵ Byrne, 'Truth in Fiction', 24.

3. Against Implicitism

The central premise of my argument against implicitism involves the notion of *works canonically related to a given work*, or *canonical relatives* of a given work for short. The defining feature of canonical relatives is that they have the power to settle the truth values of certain propositions in the fictions to which they're related.⁶ Given a work *W*, the most familiar example of a canonical relative of *W* is a sequel (or 'prequel') to *W*, though some works might have canonical relatives of other sorts.

Do canonical relatives, as I've defined them, actually exist? It seems they do.⁷ Here's an example.

In J.R.R. Tolkien's *The Hobbit*, Bilbo Baggins acquires a strange ring with the power to turn him invisible. Although Bilbo's ring plays an important role in the story of *The Hobbit*, we aren't told very much about it in that novel; we know that it previously belonged to a creature called Gollum, but much else about its nature and origin remains mysterious. In the *Lord of the*

⁶ What are the necessary and sufficient conditions for some work to count as a canonical relative of a given work? In general, I don't know. But for an enlightening recent discussion of a special case of the problem, see Roy T. Cook, 'Canonicity and Normativity in Massive, Serialized, Collaborative Fiction', *Journal of Aesthetics and Art Criticism* 71 (2013), 271-276.

⁷ Incidentally, David Lewis agrees: 'I've spoken of Conan Doyle's Holmes stories; but many other authors also have written Holmes stories. These would have little point without inter-fictional carry-over' ('Truth in Fiction', 45).

Rings stories and other writings, however, Tolkien went on to describe the career of Bilbo's ring in great detail. In those writings, we learn that Bilbo's ring is in fact the 'One Ring', an ancient artifact created by a powerful supernatural being. We learn how Gollum came to possess the ring, why it provoked such unusual behavior in its owners, and so on.

What I want to emphasize in this example is the following: once Tolkien had written *The Lord of the Rings*, it was thereafter true that, in the story of *The Hobbit*, Bilbo finds the One Ring. More generally, all the assertions made by Tolkien about the One Ring in his other writings came to be true of the object found by Bilbo in the original story. And this example is in no way unusual. Indeed, it's arguably part of what it is for a work W_2 to be a sequel (say) to a work W_1 that W_2 settles the truth values of some propositions in the fiction F_1 associated with W_1 . At the very least, if some event e of F_2 takes place after the events of F_1 (as is typical of sequels), then W_2 makes it true in F_1 that e will occur in the future. It therefore hardly seems possible to understand the relationship between a fiction and its sequels without grasping that what's true in the latter partially determines what's true in the former.

In general, there's practically no telling which propositions may be made true in a given story by its canonical relatives. There's certainly no reason why surprising revelations might not be made in a sequel, for example, that couldn't have been predicted on the basis of the original story. We might learn that a setting that was made to seem alien or fantastical is in fact Earth in a distant (or not-so-distant) future. Or we could discover that 'it was all a dream'. As a rule, there's no reason to think that writers of fiction won't withhold important information in an earlier story, only to reveal it in a subsequent work. That is, there's no reason to think that writers will or

should always obey Gricean maxims in the course of their storytelling. This should be no surprise, since it's often necessary to violate such maxims in order to achieve common dramatic and aesthetic effects.

While it's perhaps unusual for canonical relatives to surprise us in truly bizarre or radical ways, it's certainly not inconceivable that this should happen, and we'd have no difficulty understanding it if it did. For example, it's logically possible that Tolstoy, in a sequel to *War and Peace*, revealed that the marriage of Pierre and Natasha was orchestrated by conspirators from the Andromeda galaxy. If Tolstoy had actually done this, it would thereafter have been true in *War and Peace* that the principal romance was brought about by an intergalactic plot. While we might be dubious about the literary merits of Tolstoy's decision, we surely wouldn't find it objectionable on purely logical grounds. *War and Peace* and its hypothetical sequel would, I take it, comprise a perfectly consistent and intelligible story. I believe the point holds with generality: any proposition not explicitly false in a given fiction is made true by some possible canonical relative of the fiction.

Having established this, the conclusion I promised is only a short step away. If it could be, or could have been, made true in the Holmes stories that Holmes has a third nostril, then it can't (now, actually) be true in the stories that Holmes lacks a third nostril, on pain of contradiction. To see this, suppose on the contrary that 'Holmes lacks a third nostril' is true in the stories. Then, if Conan Doyle had given Holmes an extra nostril in a further story, we would apparently have to judge the result inconsistent or unintelligible, since it would then be both (implicitly) true in the stories that Holmes lacked a third nostril and also (explicitly) false that he

did. But this supposition is wrong. We surely wouldn't, in fact, find such a development inconsistent or unintelligible. (We might well find it silly, incongruous, or in bad taste, but that's another question.) Therefore, 'Holmes lacks a third nostril' can't have been true in the stories in the first place. Similarly, if it could have been made true in *War and Peace* that an alien intelligence was at work behind the scenes, then this proposition couldn't (now, actually) be false in *War and Peace*. And so on.

The main observations I've been making may be condensed into the following thesis. Let W_1 be a work and F the associated fiction. Let p be any proposition that's neither explicitly true nor explicitly false in W_1 . Then:

- (1) There exists a possible canonical relative W_2 of W_1 such that W_2 makes p false in F , and the resulting story (in particular, the set $\{\sim p\} \cup F$) is logically consistent.⁸

I hope that (1) is uncontroversial, or as uncontroversial as a premise in a philosophical argument can be; as I've said already, some such principle seems to me indispensable for understanding the relationship between a work of fiction and its sequels. Moreover, it's trivially true that

- (2) If the set $\{\sim p\} \cup F$ is logically consistent, then $p \notin F$.

⁸ Here I'm making use of the identification of a fiction with the set of propositions that are true in the fiction.

So, by (1) and (2),

(3) p is not true in F . In particular, p is not *implicitly* true in F .

Since the argument concerns an arbitrary work and an arbitrary proposition, this last statement amounts to the claim that there are no implicit truths in fictions. Hence implicitism is false.

If I'm right about this, then we need to start considering alternative accounts of truth in fiction—that is, the various possible versions of explicitism. The prevailing wisdom has it that explicitism is a hopelessly bad theory, so this might seem like an ill-fated undertaking. But I think this is wrong. As I hope to show in the next section, explicitism survives a variety of common and initially plausible objections, and in fact it comes out looking rather attractive in many ways.

4. Prospects for an Explicitist Theory of Truth in Fiction

Explicitism, as I've defined it, is the thesis that the set of true propositions in a given fiction is a subset of the set of explicit propositions in that fiction. Even if the above arguments establish the truth of explicitism, they don't by themselves point the way toward a fully satisfying theory of truth in fiction. For explicitism in this general form doesn't give a *criterion* for deciding whether an arbitrary proposition is true or not in an arbitrary fiction. We may be convinced that the truths in a certain fiction are some subset of the explicit propositions, but how this subset is to be identified remains to be seen.

Alas, I don't have a proper theory of this sort to offer. A few observations, however, can readily be made about the prospects for such a theory.

First, we won't want to say that *every* explicit proposition in each work is true in the associated fiction. For characters' thoughts and speech often take the form of explicit statements, and characters often lie or get things wrong. Second, we won't even want to say that every explicit statement made "in the narrator's voice" expresses a true proposition in the relevant fiction. For narrators are sometimes unreliable, even when they don't appear as characters in the usual sense.

It may well be that, in the great majority of works of fiction, the narrator may appropriately be viewed as omniscient and sincere, so that everything said 'in the narrator's voice' can be taken as true in the story. Even in the relatively few works where this agreeable situation breaks down, it may be that it's often easy to tell which of the narrator's assertions we should take as true and which we should reject. Unfortunately, however, even this level of convenience is not always available to the reader. There's a small but significant body of literature involving narrators who are known (or reasonably suspected) to be unreliable, but whose assertions can't easily be sorted into the true and the false or dubious. (Consider Nabokov's *Pale Fire*, for example, a key feature of which is the massive uncertainty surrounding the narrator's version of events.) In view of works such as these, it's doubtful that there exists any simple algorithm for determining whether a given explicit proposition is true in a given fiction.

Although I'm stopping short of offering a 'decision procedure' for truth in fiction, the conclusion I've argued for is nevertheless quite strong. If I'm right, then the theory of truth in fiction needs to be largely rebuilt upon new foundations. But we haven't yet reached the end of the story. In the next couple sections, I anticipate and try to respond to some concerns about the plausibility of explicitism and the arguments I've relied on.

5. Is Explicitism Perverse?

Implicitism, as its defenders have pointed out, has considerable intuitive appeal. Surely, one wants to say, there's something natural and appropriate about thinking that Holmes wears underpants, or that he doesn't have a third nostril. To remain agnostic about such things is to be badly confused about how to understand works of fiction. If explicitism invites us to engage in such bizarre imaginative behavior, then that just shows there's something wrong with the theory: the proper inferential move here is not a modus ponens yielding the nonexistence of implicit truths in fiction, but rather a modus tollens yielding the denial of explicitism.

In my view, this line of thought is largely correct, although the falsity of explicitism isn't the proper moral to draw from it. There is indeed something appropriate about the implicitist's mode of imaginative engagement with fiction; someone really is making an important kind of mistake if she earnestly refuses to imagine Holmes having two nostrils, or if she claims to have no idea whether Ophelia is a human being.

The explanation for this, in my view, is not that implicitism is actually correct. On the contrary, I stand by the claim that 'Holmes has two nostrils' is not true in the Holmes stories, in

the proper and strict sense of 'true'. (Of course it's not false in the stories either; it simply lacks a truth value.) The correct explanation is rather that it's *reasonable for the reader to imagine*, or to *tentatively suppose*, that Holmes has two nostrils. The reader is entitled to imagine this because, for instance, that's the picture of Holmes that Conan Doyle presumably had in mind, and it's the one that Conan Doyle presumably wished for his readers to have in mind.

There's a familiar kind of 'cooperative principle' at work here, imposed by the limitations of the medium and of human beings. The authors of fictions want their readers to take for granted various propositions that it would be impractical to make explicit in the story. The readers are aware of this, so they cooperatively supply whichever details they think the author probably intended.

This system is elegant and effective. Without some such coordination between author and audience, fiction as we know it could hardly exist. But the system is no foolproof guide to truth in fiction, properly so called. For a reader might have every reason to believe that the author of a given work intends for her to think that p is true in the story, and yet p might later turn out to be false in the story.

It doesn't take a sophisticated argument to see this. It suffices just to note that, if 'The reader is permitted (or obliged) to imagine that p is true in the story' entailed ' p is actually true in the story', it would follow that authors could never surprise readers by temporarily encouraging them to make some innocuous-seeming assumption, only to reveal the wrongness of the assumption at some suitably unexpected juncture. But of course authors can do this, and they

do it routinely. So it would be a mistake to confuse what is true in a story with what we might or ought to cooperatively imagine to be true.

I conclude that explicitism can safely accommodate implicitist-style intuitions, or something near enough: it can accommodate the thought, at least, that it's in some sense appropriate for readers to act as though there are implicit truths of fiction. Moreover, accepting explicitism doesn't mean adopting a bizarre or tendentious attitude toward fiction. An explicitist can agree that one ought to imagine Holmes having two nostrils (until one has good reason to do otherwise). To deny that such propositions are true in the pertinent stories is not to hold that it's necessarily inappropriate to act in certain respects as if they were true.

6. Genre and Canonical Relatives

Let me now move on to another issue. Perhaps the most attractive direct response to my argument involves the rejection of (1), which essentially claims that any work could have any sort of canonical relative whatsoever. Many people will be skeptical of this claim. In particular, it's natural to think that the *genre* of a work places some (often rather strong) constraints on what kinds of sequels it could have.

Consider the bizarre sequel to *War and Peace* described in section 3, in which Pierre and Natasha's marriage is revealed to have been planned by alien conspirators. Presumably nobody would deny that Tolstoy might have written such a work, as a matter of bare logical possibility. But if he had written it, could it really have been a genuine sequel—i.e., a genuine canonical relative—to *War and Peace*? Some will be inclined to say no, on the grounds that *War and*

Peace is a work of realistic historical fiction, and that the norms of that genre forbid such strange flights of fancy. On this view, Tolstoy might have written *some* genuine sequel to *War and Peace*—a realist novel about the role of Pierre and Nikolenka Volkonsky in the Decembrist uprising, maybe—but not a sequel involving outlandish intergalactic plots.

The general thrust of this line of thought, then, is to deny that any work could have just any type of canonical relative. The set of possibilities is circumscribed by the genre to which the original work belongs. Hence (1) is false.

I don't think this argument succeeds, however. First and most obviously, it simply isn't the case that canonical relatives have to be works of the same genre as the original. On the contrary, one frequently sees relatively mild genre shifts within a set of canonically related works, and even more drastic shifts are fairly common. For example, the first two books of Jeff Lindsay's popular *Dexter* series are naturalistic crime thrillers. But the third book, *Dexter in the Dark*, introduces the pagan god Moloch as a major character and plot element. Despite the sudden and radical shift from naturalism to supernaturalism, there's no temptation to rule *Dexter in the Dark* noncanonical (let alone knowable *a priori* to be noncanonical). So it would be wrong to suppose that the genre of a given work determines the genre of the work's possible canonical relatives in any straightforward way.

In any case, there's a second problem with the genre objection. For it assumes falsely that genre is an *intrinsic property* of a work. In general, however, membership in a genre depends not only on the features of the work itself, but also on contingent facts about the work's actual

canonical relatives, if it has any. For one can certainly imagine a genre of *multistylistic serial works*, which includes only series with two or more canonically related sub-works, and whose distinguishing feature is the occurrence of profound shifts in content and manner from one sub-work to another. Perhaps such a genre has been recognized already; at any rate, it seems obviously *possible* that works in such a genre could exist. Moreover, one clearly couldn't tell by examining a single work in isolation whether it belonged (as a sub-work in a larger collection) to the genre of multistylistic serials.

What's the point of introducing this notion? I want to suggest that, given any work *W*, whatever *W*'s *actual* genre may be, *W* could (as a matter of logical possibility) have been part of a multistylistic serial. If *W* had been such a work, it would have had canonical relatives of arbitrarily strange and incongruous kinds.

I've thus made a 'softer' and a 'harder' argument against the key premise of the genre objection. The softer argument points out that the genre of a work doesn't, as a matter of observable fact, place substantive constraints on the genres of its canonical relatives. The harder argument suggests that a work's actual genre *couldn't possibly* entail anything about the contents of its possible canonical relatives, since any work could have belonged to a genre in which it had canonical relatives of any sort.

7. A Harder Problem for Explicitism

I've gone to some trouble in the last couple sections to show that explicitism is a credible view, not to be dismissed out of hand as obviously false. I think these arguments are important ones to

get on the table, and I hope to have opened up some space for explicitism in future debates.

Unfortunately, though, my discussion has to end on a less positive note. There remains a serious worry about explicitism that I don't know how to answer, and I think the best response may in fact be to give up on (any reasonably conservative version of) the view. As it turns out, this doesn't mean a return to implicitism, but something more like a rejection of the very notion of truth in fiction.

Let me start with some setup. I previously acknowledged the implausibility of 'naïve' explicitism: that is, the view that *every* explicit proposition in a given work is true in the associated fiction. Since narrators can be unreliable and characters deceitful or misinformed, it won't do to simply believe everything we read. Hence we need a more sophisticated version of explicitism. Such a view must hold that, in general, only *some* of the relevant explicit propositions are true in the associated fiction. I left it open how this subset was to be identified, noting only the correct procedure is unlikely to be simple.

Here's an argument that seems to show that the task is not only difficult, but impossible in principle. Choose any explicit proposition p in any work W . Let p be as innocuous, or as harmoniously integrated with the larger story, or as crucially important to that story as you want. For any such explicit p , isn't there a possible canonical relative of W that makes p false in the story (such that the resulting fiction isn't unintelligible or inconsistent)? An argument very much like the one offered in section 3 seems to show that there must be. For instance, a sequel might have revealed that the narrator of the earlier work was deluded, so that even among the previous narrator's most unremarkable or seemingly trustworthy reports, some were entirely mistaken.

We might even discover that large portions of the original narrative were hallucinated or fabricated. I know of no way to establish the impossibility of any particular such development. So, by the same reasoning as before, it seems that no particular explicit proposition can be determinately true in any given work. Thus explicitism suffers from much the same type of flaw as implicitism.

What should the (would-be) explicitist say about this argument? Two general types of reply are possible. On the one hand, one could deny the conclusion and try to find some problem with the argument. This would presumably involve finding some important difference between explicit and non-explicit propositions such that the reasoning holds for one type of proposition but not the other. The second possible reply involves biting the bullet and accepting the nihilistic result that there are no truths in fictions (or no nontrivial such truths, anyway). I'll consider both strategies in turn.

As for the first strategy, I can't see any obvious difference between explicit and non-explicit propositions that wholly solves the explicitist's problem. That is, I don't know of a way to specify a rule that carves out an interesting class of explicit propositions that are immune from revision by possible canonical relatives, in the sense described above.

That's not to say that the explicitist has nothing at all to say for herself, though. There are some asymmetries between the two situations that leave explicitism looking somewhat better off than implicitism.

For instance, it doesn't seem possible for *every* explicit proposition in a given work to be made false (at one time) by some possible canonical relative. For in virtue of what would the two works then count as canonically related? Whether or not the canonical relative relation holds between a pair of works, after all, is determined in large part by the presence of significant overlap or agreement between the contents of the two works. At the very least, some degree of *compatibility* between the contents is surely required. And we have neither agreement nor compatibility in the case where W_2 asserts the falsity of absolutely everything asserted in W_1 .

So a canonical relative of some work W can't make every explicit proposition in W false, on pain of failing to meet a necessary condition on canonical relatives. This is interesting because there doesn't seem to be any similar restriction involving non-explicit propositions. There's nothing obviously problematic about W 's having a canonical relative that makes any number of non-explicit propositions false (or, alternately, true) in the story associated with W . A follow-up to *Moby-Dick*, for instance, might have agreed with everything stated explicitly in the original novel, but denied arbitrarily many propositions that the original didn't explicitly deal with ('Ishmael is allergic to latex', 'New Zealand suffered a butter shortage in 1851', etc.). There's no reason why any number of such denials should prevent one work from being canonically related to another.

To summarize, if W_2 is to be a canonical relative of W_1 , then there must be some substantial degree of compatibility between what's true in the fiction associated with W_2 and what's explicitly stated in W_1 . But there needn't be any such compatibility between what's true in W_2 's fiction and what *isn't* explicit in W_1 . This may be an encouraging result for the

explicitist. Unfortunately, though, it isn't enough to solve the problem. For even if it's true that no canonical relative of W can make *every* explicit proposition in W false (all at once), this doesn't imply that any *particular* explicit proposition in W is immune from such falsification. Nor does there seem to be any direct path from the first claim to the second. So the problem persists. Given some work W , we wanted to find a specific set of explicit propositions that are determinately true in the fiction associated with W , hence necessarily also true in any canonical relative of W . But we still don't know how to do this.

Perhaps someone wiser can solve the problem on behalf of explicitism. (Of course, any plausible solution for the explicitist's version of the problem may also be a solution for the implicitist's version, and vice versa. Any attempt to offer a solution as an argument for the superiority of one view over the other should take this into account.) But let me set the issue aside for now to discuss the other possible response to the problem.

The other response, as I said, involves biting the bullet and accepting the conclusion of the argument—that is, the claim that there aren't any (nontrivial) explicit truths in fictions.⁹ If there aren't any explicit truths, then presumably there aren't any implicit truths either, so it seems reasonable to call this view *nihilism* about truth in fiction. (It's perhaps worth noting that

⁹ 'Nontrivial' in the sense of 'not logically true'. Whether or not you should accept this qualifier depends on your views about consistency and fiction. If you think that fictions can't contain contradictions, then you should think that an explicit tautology in a work is determinately true in the work, since no possible canonical relative can make it false. If you think that contradictory fictions are possible, then there's no reason to accept this. I mostly ignore this complication in the rest of the section.

nihilism isn't incompatible with explicitism, as it's been defined here. On the contrary, nihilism is a version of explicitism, since the nihilist claim 'There are no truths in fiction' implies the explicitist claim 'All truths in fiction are explicit truths'. But of course nihilism is a highly aberrant version of explicitism, so it makes sense to treat it as a largely distinct view.)

On the face of it, nihilism seems even more incredible than ordinary explicitism. But I think a case can be made in its favor. First of all, note that nihilism needn't be as revisionary or bizarre as it might at first seem. Like other kinds of explicitists, nihilists can consistently hold that we're permitted (or obligated) to imagine or suppose certain things when we read works of fiction, even though the contents of these suppositions aren't determinately true in the corresponding story. (This was more or less the argument of section 5.) So nihilism needn't require us to massively revise our ordinary ways of engaging with fiction.

But the nihilist isn't limited to such purely defensive maneuvers. I think nihilism can also tell a reasonable story about what's wrong with the notion of truth in fiction. To the extent that we find this diagnosis plausible, we have additional reason to take nihilism seriously.

In general, the nihilist might argue, the notion of truth according to a particular source involves something like the following idea. First we have a source S of type T —a work of fiction, a personal testimony, a mathematical theory, or whatever. (The source can be represented as a set of sentences.) Second, we have a consequence relation \vdash_T for sources of type T . The consequence relation tells us how to determine what's true according to our source: a proposition p is true according to S if and only if $S \vdash_T p$.

The properties of the consequence relation vary with the type of source in question. For instance, if the source is a mathematical or scientific theory, then the familiar consequence relation \vdash of classical logic will typically be called for. (So, the claim ‘According to number theory, it’s true that there are infinitely many primes’ amounts to the claim ‘**Number Theory** \vdash “There are infinitely many primes”’.) If the source is a piece of testimony in English, then we’ll presumably want a consequence relation \vdash_E that encodes semantic facts about English in addition to purely formal or logical facts. (If Donna said ‘Steven is my nephew’, then it’s true according to her testimony that Steven is the son of one of Donna’s siblings, since ‘Steven is my nephew’ \vdash_E ‘Steven is the son of one of my siblings’.)

In order to have a reasonable notion of truth according to a source, the consequence relation appropriate for that type of source ought to have certain formal properties. One of these properties is arguably *monotonicity*. (A consequence relation is monotonic when consequences can’t be lost by adding more information to the premise set. In other words, for all sets of sentences S and S' of type T , and for all propositions p , the relation \vdash_T is monotonic iff $S \vdash_T p$ implies $S \cup S' \vdash_T p$.) The importance of monotonicity lies in the fact that monotonic consequence relations represent secure and revision-proof forms of inference, while non-monotonic consequence relations represent defeasible or provisional forms.

This distinction matters because the notion of truth according to a source presumably involves what’s definitively ratified by the source, not merely what the source tentatively or conditionally suggests. If I say ‘I like eating pears, there’s a pear here now, and I’m hungry’, for

instance, it's clearly not *true* according to my testimony that I'll soon eat the pear, even if a listener is defeasibly warranted in supposing that I'll do so on the basis of the testimony. (The listener's warrant might be defeated by the additional information that the pear is moldy, or that I'm fasting today, for example.) Such considerations suggest that only a monotonic consequence relation can give rise to a proper notion of truth according to a source.¹⁰ When monotonicity fails, we can at best talk about defeasible implication (and the non-truthlike status it bestows).

The moral, the nihilist might argue, is the following. In domains governed by monotonic consequence relations (e.g. mathematics), it makes sense to talk about truth according to a source (e.g. a mathematical theory), since we have a way to specify the information that's definitively and unconditionally nailed down by the source in question. But this isn't the case with fiction. As the previous arguments have shown, there's no reasonable (nontrivial) consequence relation that can do the necessary work here. We have at best 'defeasible supposition in fiction' rather than truth in fiction, and the logic of this type of supposition is thoroughly non-monotonic: whenever we have $W_1 \vdash_F p$, there always exists a possible work W_2 such that $W_1 \cup W_2 \not\vdash_F p$.

¹⁰ To be more precise, what's needed to justify any *particular* claim of the form '*p* is true in *F*' is just that the particular inference involved satisfy the monotonicity condition. (It's possible for a consequence relation that's non-monotonic in general to exhibit monotonic behavior in certain situations.) This qualification doesn't matter much for present purposes, though, since the arguments given above (if sound) show that any reasonable consequence relation for fiction behaves monotonically only in trivial cases, if at all.

Still, the nihilist needn't think that the debate about truth in fiction has achieved nothing worthwhile. Presumably the properties of \vdash_F (and the corresponding 'defeasibly supposable' status) are well worth investigating, and it's natural to view much of the literature so far as doing precisely this. The mistake was just to think that there's something called 'truth according to a fiction', analogous to 'truth according to a mathematical theory' and similar notions, involving a straightforward monotonic relationship between the contents of literary works and those of the associated fictions.

Is nihilism the right view about truth in fiction, or can some stronger version of explicitism (or even implicitism) be defended? I confess to having certain nihilist leanings, but more work needs to be done before the issue can be settled.

Though there's certainly more to say about many of the issues I've raised, I lack the space to venture further here. If I've done nothing else, I hope to have at least persuaded other philosophers that there's an argument worth having, and that matters are much less straightforward than has usually been assumed. I believe it's long since time that implicitism's age of innocence came to an end.¹¹

¹¹ Many people talked with me about this paper, and I'm happy to be able to thank them. My biggest debt is to Mahrud Almotahari, who read several drafts of the paper and has given me much wise counsel. I also got good advice and valuable comments from Walter Edelberg, John Holbo, Neil Van Leeuwen, Ira Newman, Mallory Webber, Lauren Woomeer, and two anonymous referees for the *BJA*, as well as from audiences at UIC and at the 2013 Annual Meeting of the American Society for Aesthetics. Finally, I thank Michael Forest and Walter Edelberg for introducing me to aesthetics, and more importantly for showing me how I wanted to do and teach philosophy.