

A Consistent Reading of “Sylvan’s Box”

Draft. Final version to appear in *Philosophical Quarterly*.

Abstract: This paper argues that Graham Priest’s story *Sylvan’s Box* has an attractive, consistent reading. Priest’s hope to use that story as an example of a non-trivial “essentially inconsistent” story is thus threatened. The paper then makes some observations about the role *Sylvan’s Box* might play in a theory of unreliable narrators.

If there are fictions that depict logically impossible goings-on, then it seems we have trouble for a range of philosophical views that have seemed tempting. The attempt to say that the limits of possibility are the limits of representation should be abandoned. Some versions of the view that something is possible if and only if it is conceivable are in trouble if we conceive of things that are the paradigm of impossible. If there are impossible fictions, then it suggests that there are impossible worlds that these fictions describe (at least in the sense that there “are” possible worlds that more mundane representations describe). If there are fictions that represent some logically impossible things, but do not thereby represent all impossibilities, that would suggest that “what ifs” do not treat all impossibilities alike: and so theories of the conditional that make all conditionals with impossible antecedents true, for example, or make them all false, are also under indirect pressure.

It is for these reasons that Graham Priest’s story *Sylvan’s Box* (Priest 1997, 2005) is significant. In the story, Priest recounts a visit to Richard Sylvan’s house not long after Sylvan’s death. Priest reports discovering a box marked “Impossible Object”, and upon opening it discovers the box to both contain a statue and to be entirely empty. After some confusion, Priest shows the box to Nicholas Griffin, who is also stunned to see such a plainly visible contradiction. Priest speculates that it was the discovery of the (non-) contents of this box that changed Sylvan’s mind in the 1970s from believing that there were no actual contradictions to believing that there may well be some. Griffin speculates that the phenomenon might be a macroscopic, inconsistent, quantum-mechanical one. Rather than deciding to tell the world, the pair, Priest says, decide to dispose of the box in an inconsistent way - Priest drives off with it in his boot, while Griffin buries it behind the house. Priest claims that this story is an inconsistent but coherent story, and the “chunking” strategy of Lewis 1983 and others, of dividing the story up into consistent fragments and allowing that what is true in the story is a function of what is true in the consistent ones, will not work to account for what is true in the story either. For it seems that the inconsistency of the contents of the box is crucial to the story: Priest says it is “essential to the plot” in Priest 1997 p 579. It is what makes sense of the characters’ reactions. Priest draws ten “morals” from the story, about the non-triviality of some impossible fictions, the need to do inference in discovering what is true in the story that does not become trivial when the premises include a contradiction, the need for impossible worlds, and several others.

One might wonder why this story is needed to make the point that there are impossible fictions, they describe impossible situations, not everything is true according to one, and so on. After all, logically impossible and inconsistent fictions are very easy to produce.

“Mr McGee did seven logically impossible things before breakfast every morning.” “The meteorologists were astounded. It really was raining and not raining in exactly the same place, and they now knew that this was in exactly the same sense.” The value of *Sylvan’s Box*, I take it, is that it has the form of a more standard short story - the writing is literary, the interest in the characters and their situation genuine, and it engages our usual powers of literary understanding in a way that perhaps my two very short stories do not. It is less clear what is lost if we treat my micro-stories as divisible without remainder into consistent sub-fragments, and they are fragmentary enough so that even quite unintended readings might not strike us as missing the point. I could imagine someone comfortably saying that everything is true according to each of them, taken at face value. I think that would be a mistake, but I think it would do less immediate intuitive violence to our conception of what is true in a story than such a misinterpretation of *Sylvan’s Box*.

However, for *Sylvan’s Box* to be a good poster-child for inconsistent fiction, it must really be the case that there is no appealing consistent reading of the story that makes sense of the piece. Priest, of course, assures us that there is not, and that “anyone who misapplied the principle of charity to interpret the story in a consistent way, would have entirely misunderstood it” (Priest 1997 p 580), but we need not take an author’s assessment of their own work at face value. Unfortunately for the case of *Sylvan’s Box*, I think there is a quite plausible consistent reading of the story. This does not show Priest’s opponents are right, of course - even if *Sylvan’s Box* were a consistent fiction, that would not show that there are no non-trivial inconsistent ones, or that there were no impossible worlds, or any such thing. Nor do I particularly want to show any such thing - I believe in non-trivial inconsistent fictions, I have little sympathy for the “chunking” strategy, and I am committed to impossible worlds and situations. So finding a problem with the *Sylvan’s Box* example is a declaration against interest. Nevertheless, the kind of strategy to be employed is worth taking note of in the debate about inconsistent fictions. Consideration of this strategy also brings up an interesting issue in philosophical aesthetics which I will discuss at the end of this paper.

Sylvan’s Box is told in the first person, and through it we see Priest grapple with the mystery he has discovered, and together with Griffin he tries to make sense of it. The reading we are intended to get is one where Priest finds a contradictory box. But the story makes a lot of sense if, instead, we read it as a story where Priest *believes* he has found an inconsistent object. The box and its contents, somehow, impress on Priest and subsequently Griffin that there is an observable inconsistent object. It is not that Priest is unprimed, either - he is in the middle of going through Sylvan’s papers, which raise to salience the idea that there might be contradictions in the world. A story in which Priest believes he has discovered a box that is both empty and contains a statue, that it strikes Griffin that way too, and that perhaps had that effect on Routley/Sylvan years ago, captures much that is of importance about the story. It makes the psychological knock-on effects of the discovery explicable, and, provided the box makes Priest and Griffin think they can carry out an inconsistent plan with it, explains why Priest reports what he does. The narrator reports finding such a box, arranging for it to be both buried and driven away, because the narrator believes that is exactly what happened.

Of course, such a reading does not say what the box was really like. Did it have some very unusual ability to affect someone's beliefs, did it actually look inconsistent, or was the presence of the box apparently incidental and the strange experiences be entirely due to some sort of derangement on the part of Priest and Griffin? Perhaps the story is best read as silent on this question, or I suppose one might argue that one attribution to the story is better than the others. People who think it is impossible for something to even *look* inconsistent might prefer one of the other ways of filling in that part, for example. That reading might still not suit some who think it is impossible to knowingly believe a contradiction, but those people have a lot of explaining-away to do about the real live Graham Priest before we need worry about what they say about stories where people explicitly endorse contradictions.

It is, of course, entirely consistent that a pair of logicians might come to believe contradictions as a result of some experiences. (Particularly when one of them is Graham Priest!) Indeed, on an autobiographical note, the first time I read the story, when it was not yet entirely obvious it was intended as fiction, I did wonder if, perhaps, Priest really did believe he had found an inconsistent object in Sylvan's house. I wondered that only briefly, but if we want some evidence of what an audience member might believe if told the story as a non-fiction recounting, then that might be relevant. I did not wonder, on the other hand, whether Priest really did find a contradictory object - I know that cannot have happened. I think a good case can be made that the story suggests to a hearer that Priest *thought* he found a strange object much more strongly than it suggests Priest really might have found such an impossible object.

This consistent reading does not make a hash of the narration, and straightforwardly tells us a lot about the setting, and why Priest and Griffin reacted as they did (or thought they reacted as they did, if we are considering their reported inconsistent actions at the end). One downside is that it appears to directly go against some of what is explicitly in the story. For example:

At first I thought it must be a trick of the light, but more careful inspection certified that it was no illusion. The box was absolutely empty, but also had something in it. Fixed to its base was a small figurine, carved of wood (Priest 1997 p 575)

or

The box was really empty and occupied at the same time. (p 576)

And there are several other passages which indicate that the box has impossible contents. However, this is not the decisive problem that it may at first appear. We are familiar with the phenomenon of *unreliable narrators* in fiction: often we get fiction told in the first person in which we pick up that the narrator is mistaken about what is in fact going on, according to the story. This may be one of these cases. The story reads like Priest is telling us what happened, and perhaps all that the above passages force us to conclude is true in the story is that *Priest is saying that* e.g. "The box was really empty and occupied at the same time". We should probably infer further - at least, I claim, that according to the story Priest believed that is what happened - there is no evidence that according to the

story he is lying or holding out on us when giving the account. This reading does not ignore these passages, indeed they contribute essentially to our understanding of the story. It is just that, on this reading, these passages signal nothing impossible about the world of the fiction - just that Priest had some unusual experiences. And even Priest's preferred reading tells us that much is strange about the world of the fiction!

Should we, on balance, decide that this is the best reading of *Sylvan's Box*? I suspect it will depend on our other theoretical commitments. Since I think there can well be non-trivial impossible fictions, and given that Priest's preferred reading is the one that initially strikes most readers, I am happy to allow that the reading on which there is an impossible, contradictory object in the story is preferable to the one where the fictional world is consistent and the characters mistakenly believe there is an impossibility about the box. But those antecedently inclined to think that there are not impossible fictions may well want to go the other way. I am offering them an account of *Sylvan's Box* which makes sense of the narrative and lets us explain what the characters are doing and why, and which does not truck with impossibility. This story may help those who prefer "chunking" approaches to fictions that apparently describe the impossible. If Priest and Griffin's false beliefs are in common to different fragments, and the fragments differ about e.g. exactly what, if anything, is in the box, then those "chunks" together will do at least as well to capture the story as the single unreliable narrator reading. So Priest's claim that his story is "essentially inconsistent" may well have little force against those antecedently inclined to disbelieve it.

To resist Priest's *conclusions* about the existence and nature of non-trivial inconsistent fictions, though, as opposed to the particular argument that seems to be offered, an opponent of Priest might have to do a little more than offer a consistent reading of his story. For even if there is a consistent reading that makes sense, if there is *also* a non-trivial inconsistent reading, not susceptible to "chunking", then many of his points, and others one might want to make, go through: for then that *reading* seems to require, in some sense, an inconsistent fictional world to be about; it requires inference about the impossible that does not go haywire when a contradiction is found; it seems to show we can make sense of, and maybe in one sense conceive of, an impossible happening, and so on. It is a truism in reception studies that there can be more than one good way of reading a piece of fiction (or other communications, for that matter): and while one might want to take the ideology of that field with a grain of salt (e.g. many analytic philosophers would want to reject the claim, close to orthodoxy in some quarters, that there is no such thing as a "correct" versus "incorrect" reading of a piece of literature), it seems we should take multiple readings seriously. And it seems to me Priest's position could be satisfied with *Sylvan's Box* having one of several readings that have the features he advertises. I recommend that line of argument to the defenders of inconsistent fiction.

The final point to note about Priest's story is that even if it does not succeed in forcing a sensible, but inconsistent, reading, some other stories might do better. One told in the third person already makes the unreliable narrator move more difficult (though some argue that all stories have the possibility of unreliable narration, e.g. Seymour Chatman 1990 who argues that there is always an implicit narrator, and Gregory Currie 1995

argues that all stories have at least an implicit author who can, in a sense, be a source of unreliability). A story where the inconsistency has a lot of down-stream effects that are not caused by characters' inconsistent beliefs would be harder to make sense of by supposing characters had been led to an error, and there are other ways to make an unreliable narration reading harder. If it is true that *Sylvan's Box* is not a great example because of the availability of a plausible, consistent, reading, then this might tempt those on Priest's side to dig out other examples of inconsistent literary fiction (e.g. from Borges) or come up with new stories that unlike the toy micro-stories I mentioned above require genuine engagement from readers.

An Observation About Unreliable Narration

Apart from its bearing on issues of representations of impossibility, the "unreliable narrator" hypothesis applied to *Sylvan's Box* raises an interesting question more narrowly about fiction. Suppose that Priest is right that, after all, an unreliable narrator reading of *Sylvan's Box* is not the most apt. This leaves us with an interesting observation about when to read a story as having an unreliable narrator. Since a story very rarely says explicitly that the narrator is unreliable, presumably we rely on cues elsewhere - and for stories that are told entirely by a narrator, presumably they are cues in what the narrator him or herself tells us. Obviously the details of how we pick up on unreliable narration are likely to be quite involved: and after all, these stories are an *artform*, and some authors are going to find subtle ways to suggest unreliable narration is happening. But presumably the plausible first pass at a theory of how we detect unreliable narration is by one of two means. The first is that we notice incoherencies in what the narrator is telling us. If the narrator tells us that his wife is happy that he is having an affair, and later tells us his wife was angry and upset with him over the affair, we might suspect that the author was mistaken about his wife's attitude in one or other place. (Of course she might be happy about some things and angry or miserable about others, or her mood might swing between the poles - but we might be very likely to think that e.g. the narrator was engaged in wishful thinking in deciding that his wife would be happy about it, or that he took at face value her insincere claim she was happy about it.)

The second way of detecting unreliable narrators might have to do with certain sorts of intrinsic implausibility, as when, in "Flash Stockman", we are told that the narrator is "the king of every blasted thing" and capable of superhuman feats, but we instead take the narrator to be a boasting liar. (See Lewis 1983 p 279-80.) Of course, what sorts of intrinsic implausibility count is a nice question, and may be genre relative: if a narrator reporting flying rocket cars, or devastation by dragons, will not thereby be thought unreliable if we are reading science fiction or fantasy, respectively. Many stories, even of more everyday genres, concern very unlikely events: it is implausible that two sisters from the lowest ranks of the gentry will marry two of the finest gentry of England, but nobody thinks this makes *Pride and Prejudice* a candidate for unreliable narration, and nor would we even if it was told in first person, for example as the memoirs of Elizabeth Bennett. The sort of intrinsic implausibility will presumably matter, in somewhat complex ways. There is also a combination - where two things the narrator says only

cohere together if we make implausible assumptions. Consider, for example, the case of Nabokov's *Lolita*, when the pedophile narrator Hubert Humphrey tells us the girl he is sexually abusing leads him on and loves him, while what he tells us about the actions of Lolita suggest she is a terrified girl trying to escape the power of Humphrey. Of course, since psychology is contingent, it is metaphysically possible for someone who is enthusiastic about a sexual relationship and in love will act in ways we would normally take to show they were very unhappy, including trying to run away. But that's a very implausible thing to happen. More likely, we should discount the narrator's reports of Lolita's attitudes towards him.

Sylvan's Box, however, should lead us to be cautious about this first pass theory, that unreliable narration is signaled when there is incoherence in what the narrator tells us, or intrinsic implausibility, or a mixture. This is because Sylvan's Box looks like a classic case of both of these things. The account the narrator gives us of the goings-on is incoherent - he both says there is a statuette in the box, and also, at the same time, that the box is empty. This literally *cannot* be true: no box can have contents like these. One of the things the narrator tells us is the negation of the other. Furthermore, the story has very high intrinsic implausibility: the story is explicitly contradictory, and even an *ingénue* who knew little of the world should be able to know enough to know that there are no boxes like that. Nevertheless, despite the story having these features, it seems probable at least that we should not endorse an "unreliable narrator" reading of the story. So whatever it is that signals unreliability, it should be more than internal incoherence or intrinsic plausibility. Perhaps it is only internal incoherence of a certain sort that matters, or perhaps flat-out contradicting oneself is not always "incoherent" in the requisite sense. Perhaps it is nothing to do with intrinsic implausibility per se, but only implausibilities of certain to-be-specified sorts that suggest unreliable narration. Or perhaps, despite initial appearances, it is some further sort of fact about a fictional narration that signals that we should take it to be unreliable. Perhaps the unreliability needs to be seen as serving some teleological end - the story would not be as good, or would not achieve some aesthetic effect, unless we postulated the unreliability. (Part of what makes *Lolita* a great novel and not a mere well-written child molestation fantasy is that we can see how Humphrey's take on things does not match what is happening in his world.) Here is not the place to construct a better theory of unreliable narration: my purpose is rather to point out how *Sylvan's Box* is an interesting data point for such a theory.¹

Daniel Nolan
Department of Philosophy
University of Nottingham
University Park, Nottingham
NG7 2RD
United Kingdom

¹ Thanks to JC Beall, Gregory Currie, Carrie Jenkins and Graham Priest for comments.

References

Chatman, S, *Coming To Terms: The Rhetoric of Narrative in Fiction and Film* (Ithaca: Cornell University Press 1990)

Currie, G, 'Unreliability Refigured: Narrative in Literature and Film', *The Journal of Aesthetics and Art Criticism*, 53.1 (1995), pp. 19-29

Lewis, D, 'Postscripts to "Truth in Fiction"', in his 1983. *Philosophical Papers* vol 1. (Oxford: Oxford University Press, 1983), pp. 276-280

Priest, G, 'Sylvan's Box: A Short Story and Ten Morals', *Notre Dame Journal of Formal Logic* 38.4 (1997), pp. 573-582

Priest, G, *Towards Non-Being: The Logic and Metaphysics of Intentionality*. (Oxford: Oxford University Press, 2006). *Sylvan's Box* appears as an appendix to chapter 6, pp 125-133