

Impossible Fictions Part I: Lessons for Fiction

Penultimate Draft. Final Version to appear in *Philosophy Compass*. Please cite the published version if possible.

Daniel Nolan

*University of Notre Dame**

Abstract

Impossible fictions are valuable evidence both for a theory of fiction and for theories of meaning, mind and epistemology. This article focuses on what we can learn about fiction from reflecting on impossible fictions. First, different kinds of impossible fiction are considered, and the question of how much fiction is impossible is addressed. What impossible fiction contributes to our understanding of "truth in fiction" and the logic of fiction will be examined. Finally, our understanding of unreliable narrators and unreliable narration in fiction needs to accommodate stories that, on the face of it, cannot possibly be true.

1. Introduction

Fictions are full of wonderful and strange goings-on. Some of the things that happen in stories are improbable and some outright impossible. Some things that happen in fiction are not just impossible relative to our normal expectations about how the world works or how people behave: they are ruled out by the laws of nature, or fundamental metaphysical principles about objects in the world, or fictions may even contain logical impossibilities.

Philosophical reflections on impossible fictions raise a range of interesting issues. This first part of the two-part series will focus on what light impossible fictions can shed on questions in the philosophy of fiction. One kind of question is about the nature and extent of impossible fictions themselves. Section 2 will offer a classification of impossible fictions, with some remarks about specific philosophical questions that arise about each kind. Section 3 turns to insights for theories

* Correspondence: Daniel Nolan, 204 Malloy Hall, University of Notre Dame, Notre Dame IN 46656, USA. Email: dnolan2@nd.edu

of fiction more generally. It will discuss the extent to which a theory of fiction should treat impossible fictions as important data; whether impossible fictions show limitations of accounts that have been offered of what it takes for something to be true in a fiction; and what we can learn from them about the "logic of fiction". Finally, section 4 considers lessons they have for us in constructing a theory of when the narration in a fiction is "unreliable", not getting things right about the fictional world in question.

First, however, let us look at some kinds of impossibilities we find in fiction, and how they might be interestingly different. This will also put some examples on the table for the rest of the discussion, though readers will no doubt be able to think of many cases of impossible fictions besides the ones discussed.

2. Kinds of Impossibilities

A lot of what happens in fiction is impossible in some sense or other: we know it did not happen, and things do not in fact work that way. The kind of impossible fictions that have been of primary philosophical interest to date have been impossible in stronger senses. In particular, metaphysical impossibilities and logical impossibilities have received the most attention. Let us distinguish some kinds of impossibilities that are relevant to philosophical concerns, and discuss some of their distinctive features.

2.1 Nomological Impossibility.

Nomological impossibilities violate the laws of nature. This can happen because authors are not trying to make events they describe compatible with the laws of nature, as in many stories of magic, or technological wonders in soft science-fiction; or it may be because an author has tried to limit their fiction to what is allowed by our laws of nature but has been in error about what the laws allow, as in outdated "hard" science fiction, or supernatural stories that reflect errors about what can genuinely happen. One case of science fiction superseded by theoretical discoveries is Hugo Gernsback's 1911 story *Ralph 124C 41+*, which includes an "Ultra Generator" depleting the local ether field (p 35) while causing "an enormous etheric disturbance" many miles away (p 38) (Gernsback 1950).

Some philosophers maintain theories of the laws of nature according to which any violation of those laws is in fact metaphysically impossible (see e.g. Shoemaker 1980, Swoyer 1982.) A central motivation for this stance is the thought that how physical properties behave is *essential* to them: it would just not be *mass* if it did not gravitationally attract other masses, or *negative charge* if it did not repel negative charge and attract positive charge. One way some in this camp treat the apparent possibility of law violations is to treat them as representations of *alternative sets of properties*: so a story where an "anti-gravity device" switches off the attraction between masses might just be a story where we are not dealing with *mass* or *gravity* at all, but a story of a possible scenario where some alternative to mass is present instead.

Whatever the merits of this move in defusing intuitions about the metaphysical possibility of violating our laws of nature, or in handling counterfactual suppositions about e.g. gravity obeying an inverse cube law, or there being perfectly rigid rods or frictionless surfaces, it is not clear we could adopt it without violence to the contents of the stories we tell. If the author tells us an anti-gravity drive negates gravitational attraction between masses, or a faster-than-light drive sends ships through space faster than light and then back to slower-than-light at their destination, then on the face of it the author's say-so about these matters is good enough: they are stories about *mass* and *gravity* and the speed of *light*, not stories about something else, in a way that might be hidden to both authors and readers. Perhaps it would be better for those who think the laws of nature are metaphysically necessary to take these stories as genuinely representing metaphysical impossibilities.

2.2 Metaphysical Impossibilities

A common view among philosophers is that there is a stronger grade of impossibility than nomological impossibility: that some things are *metaphysically impossible*. Philosophers disagree about what it takes to be metaphysically impossible¹, but there is some (though not complete) agreement on examples. Examples of apparent metaphysical impossibilities include

¹ See Nolan 2011 for a discussion. It is possible that some of these apparent disagreements are the result of philosophers talking past each other: that is, they are just using the expression "metaphysical possibility" and its cousins in different ways. A lot of the time, though, I suspect philosophers have managed to latch on to a kind of possibility and are offering rival theories of *it*.

human beings also being prime numbers; the colour red being a sports drink; five-sided squares; and an object existing while failing to be self-identical. One motivating thought about metaphysical impossibilities is that they go against the fundamental natures of things: prime numbers are just *not the sort of thing that can be* human beings, or vice versa.

Here are some plausible examples of metaphysical impossibilities in fiction. In Lewis Carroll's *Alice in Wonderland*, the Cheshire Cat vanishes into thin air. But he does this gradually, finally leaving his smile when all the rest of him is gone. On one natural interpretation of the text, this is supposed to happen even though the Cat's teeth and lips have gone. But plausibly it is impossible to have a free-standing smile: a smile must be some kind of arrangement of a mouth. (Cameron 2015 also discusses this example.)

In Jorge Luis Borges's *The Circular Ruins*, a wizard turns out to be only a feature of his own dream. Whatever the ontology of (merely) dreamt characters, plausibly no dream character can literally dream itself into existence. Another example from Borges is *The Aleph*, where an object is found that contains the whole cosmos as parts, but which is lodged in a staircase in a cellar on Earth. It is thus a part of something which is a proper part of it, which is impossible in standard theories of parts and wholes. In some time travel stories and movies, things are a certain way at a given time, and also not that way at that very same time (see the beginning and ends of the movie *Back to the Future*). In fictions that use the trope of personification, abstract entities like death or duty can ride horses, have hands, or speak authoritatively. But plausibly abstract objects like these cannot even possibly have these sorts of human features (see Nolan 2015). Though see Bourne and Caddick Bourne 2018 for an interpretation of personification consistent with treating fictions as all depicting possibilities.

One genre in which metaphysical impossibilities abound is the genre of *meta-fiction*.

Metafictional works consciously use the fact that a work is a fiction in a number of ways: the characters might escape the story and harass the author, or address the audience directly, or hop from one fictional world to another. Some works in this genre draw attention to the artificial or created nature of a fiction in ways that might be unusual but are not impossible. But some

arguably introduce metaphysical impossibilities: whatever fictional characters are, they cannot come to life and berate their authors.

Some relatively benign metafictional techniques raise puzzles about what is going on in a story that might be best resolved by positing metaphysical impossibilities. In Shakespeare's *Henry V*, the Chorus figure appears on stage to address the audience, sometimes talking as if he is a figure in the events described (using the present tense to talk about when the King of France is attempting to suborn Henry's vassals, for example); sometimes addressing the audience about the play; and sometimes talking as if Henry and his nobles are on a stage, and e.g. the battle of Agincourt only involved a few men instead of thousands. When the Chorus says of the author of the play that he has been "[i]n little room confining mighty men", what is going on? In real life Shakespeare did not confine Henry to a small room, and in the story (bar the chorus's remarks) no author confines Henry to a small room. There are a range of interpretive options here, but one is that, at least when the Chorus speaks, it is fictional that Henry strides around a small stage and fights Agincourt with a handful of supporters on a stage in England (or wherever the play is performed), *and also thereby* that Henry fights the battle on a field in France with a host of thousands. Perhaps it is not possible for anything to both be a wooden stage in England and also a field in France, but we know how to play along with the Chorus.

2.3 Mathematical Impossibilities

A distinctive kind of impossibility that occurs in some stories is the violation of mathematical principles. One example of a mathematically impossible fiction is "The Tower of Goldbach" in Gendler 2000, where 12 is not the sum of seven and five, and another is "Division by Zero" in Chiang 2002, where division by zero is discovered to be well-defined and yields different answers for different integers. One of the earliest works of science-fiction, Louis Geoffroy's *Napoléon Apocryphe* (Geoffroy 1841), contains a mathematical fiction in passing among many other marvels. In the advanced world where Napoleon triumphs and spreads French culture throughout the world, squaring the circle is achieved (p 329). While in the story squaring the circle was "long thought impossible" ("longtemps crue impossible"), in the actual world the result *is* mathematically impossible.

It is not usual to think that mathematical impossibility is a special *grade* of impossibility: usually, mathematical impossibilities are thought to be one kind of metaphysical impossibility, or perhaps one kind of logical impossibility. Mathematical impossibilities are interesting in their own right because of a few distinctive features they have. One is that it is often easy to grasp what the scenario involves as far as the physical world and people's behaviour in the story are concerned: often in these stories, but not always, what goes on in the concrete world is not radically different just because some mathematical result turns out differently. The only kinds of non-mathematical differences tend to be in which mathematicians become famous or how research priorities shift in mathematics departments.

Another distinctive feature of some mathematically impossible fictions is that it is easy to see how something could be mathematically impossible but not *obviously* mathematically impossible. In the squaring the circle example, this was a geometric result that mathematicians searched for in all seriousness for thousands of years, with a widely accepted proof of its impossibility not being produced until 1882. It is relatively easy to see how readers could make sense of this part of Geoffroy's story, since whether it could be done was an open mathematical possibility when he wrote. Even today, even among those who understand what squaring the circle is supposed to be, the proof that it cannot be done is not trivial to understand.

2.4 Logical Impossibilities

Some fictions appear to represent scenarios ruled out by widely-held logical principles. The most commonly discussed kind of impossible fiction like this is *inconsistent* fiction: a fiction F where *According to F, A* and also *According to F, not-A* are both correct. There are two interestingly different kinds of contradictions in fiction: one where the contradiction seems inadvertent, perhaps due to a slip-up by an author or different authors wanting to do things in their own way; and one where the inconsistency is more central to appreciating the story or understanding the world of the fiction. Since these seem to call for separate treatments, and have sometimes been treated differently in philosophical theories of inconsistent fictions, they will receive their own subsections, below.

There is some discussion of fiction that violates logical laws other than the law of non-contradiction. Routley 1979, for example, suggests that there can be fictions where a conjunction obtains but neither of its conjuncts do (p 11). Apparently logically impossible fictions other than inconsistent and incomplete fictions have received comparatively little attention, though one recent exception is Wildman 2019's discussion of *empty fictions*, that have no propositions true at all according to them. If there are empty fictions, presumably not even logical theorems are true according to them. On the usual conception of logical theorems they are all logically necessary, so a fictional world where they failed would be logically impossible.

Fictions are often taken to be *incomplete*: that is, for some propositions A , neither A nor $\text{not-}A$ is true in the fiction. This is not normally taken to be enough for the fiction to be impossible; instead, this is treated as the fiction being a set of propositions which *could* be true at a possible world, though any possible world where they were true would be "completed" in one way or another. An interesting distinction can be drawn between different kinds of incompleteness in fiction. The first, most common, kind is one where the story could be filled in in various different ways but they are left unspecified by both the text and what we import. For example, many houses in twentieth century fiction have plumbing, but exactly where all the pipes are in fictional twentieth-century houses is usually not specified. A second interesting kind of incompleteness is where a story leaves some feature unsettled in a way that matters for understanding the story. According to Williams and Woodward (forthcoming), this often happens when something is deliberately ambiguous or puzzling, without being definitively resolved. A classic example of this is arguably the "ghost" in Shakespeare's *Julius Caesar*. Brutus clearly thinks he is being visited by Julius Caesar's ghost: but Brutus is tormented by a guilty conscience and perhaps an unlucky fate as a result of murdering his mentor Caesar. So the "ghost" may just be his delusion, especially since the main time Brutus "sees" the ghost is when Brutus is somewhere between being awake and asleep. Is the ghost real in the world of the fiction? Plausibly, Shakespeare deliberately leaves this ambiguous.

A third kind of incompleteness would be if the fiction represents some matter *as* incomplete: that is it is not just that the fiction does not say enough about a topic, but says that there is no fact of the matter about a topic. The most obvious kind of case would be a fiction that explicitly set out

to describe a scenario where there was a gap in reality (not just an empty space, but where e.g. the gods destroy all the facts about whether or not I wear a hat). But depending on what truth-value gaps there are in the real world, fictions with truth-value gaps may be far more common. On some theories of vagueness, for example, borderline cases of falling under a predicate yield cases where there are truth value gaps. For example, since "bald" is vague, as well as bald people and not-bald people there are some people about whom it is neither true nor false that they are bald. If this is enough for "represented as" incompleteness, this form of incompleteness is rampant in fiction: every time a story passes from early in the morning to the afternoon, or every time a character starts out with a full head of hair and gradually goes bald with age, etc. These are all cases where there is a vague transition with borderline cases in the middle.

The question of whether logical necessity is a grade of necessity somehow stricter than mere metaphysical necessity is a controversial one. (See Nolan 2011 for a discussion of some of the positions taken in the literature about the connection between the two.) It should not matter too much for current purposes whether logical necessity is a stricter grade of necessity (so that e.g. something can be logically possible but metaphysically impossible), or whether logical impossibilities are just a species of metaphysical impossibilities. One reason logical impossibilities are distinctive in fiction, whether or not they belong to their own grade of necessity, is because of their connection with *logic*. A natural thought is that one way to draw more out about what is going on in a fictional world besides what is explicitly said is to use logic to put together parts of the story. Stories that go logically awry suggest that this is not always a safe way to proceed.

2.4.1 Inadvertent Inconsistencies

David Lewis (Lewis 1978 p 46) offers a famous example of a surprising inconsistency in a fictional work: the case of Dr Watson's war wound. According to the *Sherlock Holmes* stories written by Arthur Conan Doyle, Watson was invalided back to Britain after the Afghan war in part due to a debilitating war injury. But the different *Sherlock Holmes* stories appear to place the injury in different places. *A Study in Scarlet* places it in his shoulder and *The Sign of Four* appears to place it in his leg, and neither suggests he suffered two such wounds. (This is perhaps not the best example, since there is a consistent reading of the stories where the leg-wound of

The Sign of Four is a second troubling wound Watson received in Afghanistan. But the general phenomenon of multi-part fictions contradicting themselves on details is certainly a real one.)

A natural way to interpret this is that the stories taken as a whole represent inconsistent things about the world of Sherlock Holmes: Watson has a single serious war wound, and it is in his shoulder (and not in his leg), *and* it is in his leg (and not in his shoulder). From one point of view it is not even surprising that fictions can be found like this. Large *non-fiction* works often have inconsistencies in them: nobody expects every page of *Wikipedia* to agree with every other about every factual claim. Even a single author trying to keep a story straight might inadvertently fill in details in clashing ways. But from another point of view, Watson's war wound raises some puzzling questions. If different *Wikipedia* articles contradict each other then at least one is wrong about the facts. But there do not seem to be "facts" behind Conan Doyle's story to settle which story is right about Watson. So perhaps both are: but Watson is supposed to be (according to the story) a fairly ordinary man, but how could an ordinary human being have a wound entirely in his shoulder and entirely in his leg?

Some of these issues can be avoided, for some purposes at least, by relativising features of Watson to stories: one story says one thing, another story says another, and in neither does he have a mysterious inconsistent wound. That leaves open the question of what is true about Watson in the stories as a whole, and puzzling questions still arise when inconsistencies creep into a single story. David Lewis suggests that even in these cases, we can think about the stories as a whole being associated with consistent "revised versions" or "fragments" modified in different plausible ways to remove the inconsistency, and make sense of what is true in the story as a whole in terms of what is true according to these versions. (Lewis 1978 pp 45-46, Lewis 1983 p 277). One way to say what is true according to the story as a whole is to take the *intersection* of what is true at the largest consistent fragments: on this approach, it is true according to the stories that Watson returned from Afghanistan with a serious wound, and it is true that e.g. it is not in the middle of his face, but it fails to be true that it is in his shoulder, and it fails to be true that it is in his leg. (Lewis 1983 p 277) Another approach Lewis suggests is to take the *union* of propositions true at all the different consistent fragments: then we get the plausible result that the stories say the wound is in his leg, and they say that the wound is in his

shoulder. According to this "method of union", a claim is true according to the fiction if it is true according to at least one of the fragments. Lewis points out an interesting feature of the method of union: while e.g. the Sherlock Holmes stories say that Watson has his serious wound in his shoulder, *and* they say that Watson's serious wound is not in his shoulder, the conjunction, that the wound is both in and not in his shoulder, is *not* true according to the stories. That conjunction is not true according to any of the consistent fragments. We have jointly contradictory claims being true according to fictions without any single contradiction being true. (In this system it does not follow from "According to the fiction *A*" and "According to the fiction not-*A*" that "According to the fiction *A* and not-*A*".) (Lewis 1983 pp 277-8)

We may wish to distinguish different kinds of inadvertent inconsistencies. The most puzzling are ones which are both relevant to the story and occur in the same work, or in works equally authoritative about "the" fictional world in question. Others can arise when one work is "canon" and another work is secondary or derivative. When the inconsistency is inadvertent, we are more inclined to think the derivative work is "wrong" about the matter in dispute. On the other hand when the conflict is deliberate, as when *Star Trek* fan fiction describes a romantic relationship between Kirk and Spock, we are more likely to retreat to version-relativity: there isn't such a relationship in the original series, but there is in the particular work of fan-fiction. Another case again arises when a series of stories are told about common characters but none are particularly authoritative. In some tellings of the Robin Hood stories, Robin marries Maid Marian in the end. In others, he does not. Are they married according to the Robin Hood stories? We might feel less pressure to give a once-and-for-all answer than in the case of questions about the world of Conan Doyle's *Sherlock Holmes* stories.

One issue arises when the explicit content of a fictional representation is consistent, but when one of the most tempting interpretations is that there is an inadvertent inconsistency. Could it ever be that we take an explicitly consistent presentation and infer an inconsistency? I think there are cases that make this plausible. Consider a continuity error in a film where a character's cigarette is half-burned through in one shot and then after a brief cut-away (e.g. a sentence of conversation) the character is holding a barely-burnt cigarette. It is metaphysically possible for things to be as they are explicitly depicted: perhaps the character has a magic re-growing

cigarette, or is a master at sleight-of-hand and swapped lit cigarettes without their conversational partner noticing. But we might prefer to interpret the scene as a whole as slightly inconsistent, in particular not telling an altogether consistent story about how fast the cigarette burned down. I suggest that for many continuity errors we should take this approach, rather than searching for in-the-story reasons why a hat suddenly changed colour or why a removed jacket is suddenly back on. Or at least this is what we should do when trying to recover the authorised or central interpretation: it can be fun of course to think of ways to consistently incorporate continuity errors into a film, TV show or even written literature. (Lewis 1978 p 46 suggests treating these sorts of errors using the same techniques he uses for inconsistent fictions, though he describes this sort of case as a kind of consistent fiction, albeit an unusual kind.)

2.4.2 Integral Inconsistencies

In other cases, an inconsistency seems crucial to the plot of a story. In Graham Priest's *Sylvan's Box* (Priest 1997), Priest discovers why his old friend Richard Sylvan had a sudden conversion to thinking there were true contradictions in our world: Priest finds a marvellous box among Sylvan's effects, a box that both contains a statue and is also completely empty. This visibly inconsistent object has a profound effect on Priest and his companion Griffin. It seems crucial to the story that the box has something inconsistent about it. Without that, it is hard to understand why Priest and Griffin are so perplexed, and why they come to see the box as a key to understanding Sylvan's earlier conversion.

Other stories can plausibly be read as requiring inconsistencies, even though the author does not draw as much attention to the fact that the goings-on are inconsistent. Many time-travel stories exhibit a "second time around" phenomenon: the very same time is one way at one stage of the story, and a different way at another stage of the story. Someone might be dead at t and also not dead, but rather alive, at t . (Most of the interesting cases involve giving away spoilers, but at the risk of giving a spoiler, this happens to Lois Lane in the 1978 *Superman: The Movie*. But hopefully you knew Superman would save the day somehow.) This is plausibly inconsistent: someone cannot be both alive and not alive at the same time. And as the stories are intended, it is often important that it be the same place and time, and not just e.g. a different branch of a

branching universe. Superman does not save the day if Lois stays dead on her initial timeline but he merely jumps away to an alternative Earth to be heroic in front of alternative-Lois.

In cases like these, we seem to lose something important to the story if we divide it into consistent sub-fictions then take the union of them, as in the Lewis "union" strategy mentioned above. In a story where Priest finds an empty box, and another story where he finds a box with a statue, we do not get an explanation of why he is so surprised. Neither a story where Lois dies in the aftermath of an explosion, nor another story where she never died, do we get a satisfying account of why Superman takes the grave step of travelling back in time to save the day, since in the movie this is motivated by the desire to bring Lois "back to life", and his plan does not work unless she is alive at the end of the proceedings. Priest 1997 p 580 argues that Lewis's "fragmenting" strategies will not work to do justice to his story *Sylvan's Box*, and Lewis seems to have agreed (Lewis 2004).

Not every case of inconsistency in a story will fall neatly into the baskets of "inadvertent" and "integral". There is nothing stopping an author deliberately inserting an inconsistency that is not otherwise needed for a story: writers of whimsical fictions, in particular, might occasionally make a point of it. An inconsistency might be vital for a story as it was told, even if it was inadvertent in the sense the author did not intend to tell that story in particular, but rather a variant. And how dispensable an inconsistency is to a fiction is presumably a matter of degree rather than an on/off matter.

2.5 Direct Reference and Names

There is one well-known philosophical motivation for thinking that most fictions could not even be possibly true. According to a range of theories of proper names—including most famously the account offered in Kripke 1980 of proper names and natural kind terms—names that in fact do not refer, and empty natural-kind terms, do not refer to any possible object. This is because names are tied to their referents and how they came to refer. If our actual words "Sherlock Holmes" or "unicorn" do not trace back to a particular person or to unicorns, then there is nothing for them to be picking out, in this world or in other possible worlds. So "unicorns dance

in the sunlight" is not even possibly true, since nothing could possibly count as a unicorn (Kripke 1980: 24, 156-158).

Some who hold this view think that names in fiction *do* refer to fictional entities (Kripke 2013, Salmon 1998, Braun 2005). But those who hold both that "fictional names" fail to refer, and also hold the view that when they fail to refer they could not refer, will likely hold various innocuous-looking claims in fiction could not be true. (Any claims about Sherlock Holmes, for example.) It is possible to go further, and to deny that sentences including fictional names are even *meaningful*. One well-known view of the meaning of names is the so-called "direct reference" view of names, according to which the meaning of a name *just is* the object it refers to. No object referred to, no meaning. Walton 1990 pp 402-3 appears to endorse such a view of fictional names. On one version of that view, the one Walton appears to have in mind, there are no propositions for the fiction to be associated with, so no way the propositions it expresses could be true. On a slight variation (discussed by Braun 2005 for the case of genuinely empty names), a sentence with a non-referring name can be associated with a *gappy proposition*, but "atomic" gappy propositions cannot be true. For Braun, <___, being-a-detective>, the proposition expressed by using an empty name put together with the predicate "is a detective", is necessarily false. (To be clear, Braun thinks that many uses of fictional names do pick out an object, the abstract fictional character. But he does think some do not pick out anything (p 611), and some or all of those uses correspond to gappy propositions.)

3. Insights into Fiction

3.1 How Widespread Are These Impossible Fictions?

Some forms of impossible fiction are frequently encountered. Nomologically impossible fictions can be found all over fairytales and soft science fiction, and even works intended to be realistic where the author has made some mistake about the scientific backdrop. Inconsistencies within a single work may be very common if e.g. "continuity errors" in films amount to inconsistencies about what a character is wearing, or how far their cigarette has burned down, or where the furniture in the background is. Inconsistencies in *bodies* of fictional works are certainly common: take the stories of King Arthur or Robin Hood *as a whole* and they will conflict on many points. Or consider J.K. Rowling's Harry Potter stories *together with* the fan-fiction they have generated,

and there will be enough inconsistencies between different works to keep a small army of tabulators busy for decades. And of course if certain kinds of direct reference theories about fictional names are correct almost every paradigm fiction will be impossible, in virtue of the names it contains.

It may be slightly more controversial how common metaphysically impossible fictions and (non-inadvertent) logically impossible fictions are. In the case of metaphysical impossibilities, if we count an object moving from one natural kind to another as impossible, or we think that trees and rivers cannot be the sorts of things that can talk, many stories of magic and transmutation will be metaphysically impossible. Stories with surreal or dream-like stretches are familiar from a wide range of time periods and cultures as well: *Alice in Wonderland* was mentioned above, but it is far from alone. Nolan 2015 argues that cases of *personification* of abstract things in fiction can give rise to metaphysical impossibilities: and personification is a trope employed by authors from Milton to Shakespeare to Wordsworth, to mention a handful from the traditional English canon. Finally, meta-fictional techniques where characters address the audience, or recognise that they are merely-fictional characters, or authors otherwise play with the fiction/reality boundary, are very common when you begin to look. Cameron 2015 argues that many metaphysical impossibilities are found in speculative fiction: see pp 40-41 for a list of examples.

Insofar as impossible fiction is a phenomenon of interest in its own right, it is widespread and enjoyed well beyond the examples produced by philosophers or others intending to produce mind-bending work. A theory of fiction should be judged just as much by how much justice it can do to impossible fictions as its ability to satisfactorily resolve puzzles about possible ones.

3.2 Truth in Fiction

When it comes to our understanding of fiction itself, impossible fictions post challenges for natural accounts of the content of those fictions. One common use of the expression "truth in fiction" among philosophers picks out the phenomenon of people taking various propositions to be true *according to* a particular story. These include things that are explicitly stated in the text of the story ("In a hole in the ground there lived a hobbit", in *The Hobbit*, for example), and things that are only conveyed implicitly (e.g. that there is a breathable atmosphere in the Shire in

The Hobbit). One challenge philosophers of fiction have taken on is to try to explain, in general, what counts as true according to a given story, and some of the main ways readers determine what is going on in the story from the text and other cues.

(There are other uses of the expression "truth in fiction": one common one outside these philosophical discussions is associated with the fact that we can often learn truths about the real world from experiencing or reflecting on the content of fictions.)

You might have thought one way we could safely move from some propositions true in a fiction to others is to look at what the first set of propositions *entails*, or what logically follows from them. Those entailments, you might think, must also be implicitly true in the relevant fiction. Impossible fictions, especially logically impossible fictions, cause trouble for this natural thought. According to classical logic, any formal contradiction (a claim of the form *A and not-A*, or pair of premises of the form *A; not-A*) entails every proposition whatsoever. If this principle is a correct logical principle, and fictions had to be "closed under entailment" (i.e. when a set *S* of propositions entail a proposition *B*, then if all of *S* are true according to a fiction, *B* is also true according to that fiction), then fictions where a contradiction is true would have to have every proposition whatsoever true according to them. But inconsistent fictions are not trivial like this.

In response to puzzles like this, philosophers have argued about *the logic of fiction*. Given a set of claims *S* true according to a fiction *F*, what else is deductively guaranteed to be true according to *F*? A set of propositions weaker than all the classical-logical consequences of *S*, since not everything is true according to *F* when a claim of the form *A and not-A* is true according to it. We also want to allow for gaps in what is true in a fiction: we should not require that one of a proposition or its negation is true in fiction. And so on for various other tempting principles. Routley 1979 pp 10-11 argues that there is *no* logic that holds across all fictions: or rather, that the strongest logic that always applies allows us to go from the set of claims *S* only to subsets of *S* itself. (Woods 2018 is also critical of the usual project of offering a logic of fiction, even though his earlier work attempting to articulate a logic of fiction was one of the starting points of the contemporary discussion of this topic.) Even Routley thinks there is a logic that applies to *almost* all fictions, though, corresponding to a weak logic in the relevant logic family (pp 13-14).

The more anarchic logically impossible fictions can be, the weaker the logic that applies to all fictions will have to be. Proudfoot 2018 argues that any constraint on truth in fiction we might come up with can be counterexampled by an ingenious enough author, making the prospects for even very weak logics of fiction look dim. Recently, other logical approaches have been proposed to make room for the fact that impossible fictions can be very badly behaved, logically speaking. Badura and Berto 2019, for example, offers a non-monotonic logic of fiction, where apparently legitimate deductions about what is true in a story might be withdrawn at later stages of reasoning as new information about the fiction and its context become available.

A number of strategies for accounting for truth in fiction have been developed that make room for impossible fictions. Routley 1979 does not offer a detailed positive proposal, but he points out that sometimes fictions obey what Woods 1974 calls a "sayso condition" on fictional truth: at least in some stories, an author saying A is part of a story is enough to make it be one. Some stories may not satisfy this sayso condition, but if authors *can* insist on a sayso condition at least in favourable contexts, all manner of impossibilities can be generated in a story just by author stipulation. (On the other hand, as will be discussed in Part II, authors including Hanley 2004 and Xhignesse 2016 argue against the proposal that author's say-so can be carried this far.) Currie 1986 offers a general account of truth in fiction that easily carries over to the impossible case. According to Currie, when interpreting a fiction we attempt to work out what would have been reasonable to infer about what an author would have believed if s/he had been sincerely asserting the claims. (Currie 1986 p 201). To work this out we use all our normal understanding of communication, along with cues such as the genre of the work. This makes it easy for impossibilities, even logical impossibilities, to be true according to a fiction, at least once we realise that ordinary believers can have beliefs that cannot possibly be true and may even be inconsistent or otherwise illogical.

I should note that there are reasons to be suspicious that even possible fictions are "closed under entailment": that is, that whenever there is a set of propositions S that are true according to a fiction A , and in fact S jointly logically entail B , then according to the fiction B . In mathematics, some very complex and abstruse theorems follow from the basic axioms of number theory, as a matter of logic. But while we might suppose the basic axioms of number theory obtain in a story

where merchants do some calculations of quantities or prices, there may be no particular reason to suppose that every theorem of number theory is true in the story, particularly if we suspect the author and intended audience had no inkling of those theorems. We might think that typical stories have limited subject matters, and are silent even about logical theorems involving radically different subjects. (Is it true according to the Sherlock Holmes stories that either there are more than one billion neutron stars or there are not? Perhaps the stories say nothing either way about this proposition, even implicitly.) Finally, as we have seen above, theorists like Lewis take some "continuity errors" to result in fictions that are possibly true as a whole but should nevertheless be treated like we treat impossible fictions. While I have suggested these fictions might be seen as implicitly inconsistent and so impossible after all, those who side with Lewis will still likely resist drawing all the inferences logically available.

3.3 Reliable and Unreliable Narrators

A final way impossible fictions may make a difference to our account of truth in fiction is that it gives us reason to revise an initially plausible account of when a fiction has an *unreliable narrator*. Some fictions, particularly those told in the first-person, are constructed in such a way that the reader is supposed to realise that the narrator's version of events does not match what "really happened" in the world of the fiction. A narrator's self-justifications may be too transparently self-serving, or the reader may have been given a glimpse into what happened through the viewpoint of another character or through some omniscient narration, or the narrator may give signs of descent into delusion or madness, or whatever. One interesting problem in the theory of fiction is to try to give a general story about when we should take a narrator to be unreliable. Narrators are rarely if ever *entirely* unreliable, so the challenge is usually to work out which parts of the presentation are to be "trusted" and which not.

It would be very difficult to characterise all the ways unreliable narration can be present in a story, or a story-on-an-interpretation. One natural suggestion is that it is a sufficient condition for a narrator's being unreliable that what the narrator says is inconsistent: if the narrator's tale contradicts itself, the narrator has undercut their own version of events. A related suggestion is that a strong piece of evidence for unreliable narration is *incoherence* in a slightly weaker sense. If the pieces do not fit together making sensible background assumptions, that is a reason to

think that, in the fiction being presented, we should not accept what the narrator says at face value.

Impossible fictions, especially inconsistent fictions, cause trouble for both of these suggestions. (See Nolan 2007 pp 671-673.) Narrators telling deliberately inconsistent tales contradict themselves, and perhaps guarantee that the story is an impossible fiction, but they do not thereby show that they cannot be "trusted" to get things right about the world of the fiction. If a continuity error means that a cigarette is almost burned down at the end of one shot and barely started a few seconds later, that should not make us suspect that there is a dark secret about the world of the movie that the surface presentation is misleading us about. That leaves the question of when we *should* postulate an unreliable narrator still to be answered of course, and there is a rich literature on unreliable narrators, particularly in literary theory.

An alternative way of thinking about the relationship between unreliable narrators and impossible fictions is to hold that inconsistencies or similar incongruities are often a sign, not that we have an impossible fiction on our hands, but that we *do* have a possible fictional world being presented to us, albeit by an unreliable narrator. Hanley 2004 pp 120, 124, 125 argues that a range of paradigm impossible fictions, including Graham Priest's *Sylvan's Box*, are possible fictions where we are presented with unreliable narrators with inconsistent beliefs. (There cannot be a box that is both empty and not-empty, but there can be someone who is deluded into thinking there is such a box!) If we wished to use inconsistency as strong evidence that a story is being presented by an unreliable narrator, that would let us avoid interpreting nearly as many fictions as impossible. As Nolan 2007 points out (p 671), it is harder to accept that there is an unreliable narrator when a story is told from a third-person perspective, particularly in an omniscient third-person way: but even in these cases we might sometimes posit potentially deceptive motives for what we are being told or shown, since even in these cases there is selection of what we are being told about, and implicit value judgements in presentation. So even in these cases we might resort to thinking there is unreliable narration when the pieces do not fit together. (See Currie 1995 for a discussion of unreliable "narrative", even for cases where there is no obvious narrator, and perhaps no non-obvious one either.)

4. Conclusion

Fictions that describe impossibilities are a familiar feature of literature, though disputes can of course be had over which fictions exactly depict impossibilities, and what sorts of impossibilities they are. Realising this has significant ramifications for our theory of fiction, including constraining what sorts of stories are plausible about truth in fiction; what a "logic of fiction" could look like; and what theory we should give of reliable and unreliable narrators.

Impossible fictions also shed light on important philosophical questions beyond the philosophy of fiction, and these will be the focus of Part II. The second part of this two-part series will also go into some detail about the options for resisting the claim that there genuinely are impossible fictions; or alternatively for resisting the thought that impossible fictions come in the variety discussed above. So readers wondering whether it is best to push back on the lessons advertised above are invited to think about the alternatives explored there.

Acknowledgments

Thanks to Sara Bernstein, Franz Berto, Ross Cameron, Jessica Collins, Hannah Kim, Roy Sorensen, Zeynep Soysal, Nathan Wildman, Richard Woodward, an anonymous referee, and audiences at the University of Rochester, Arché at the University of St Andrews, and the Slovak Academy of Sciences for feedback and discussion.

Short Biography

Daniel Nolan is McMahon-Hank Professor of Philosophy at the University of Notre Dame. He is the author of *Topics in The Philosophy of Possible Worlds* and *David Lewis*, and has published articles in a range of journals including *Noûs*, *Philosophical Studies*, *Philosophy and Phenomenological Research*, and the *Australasian Journal of Philosophy*.

References

Badura, C. and Berto, F. 2019. "Truth in Fiction, Impossible Worlds, and Belief Revision". *Australasian Journal of Philosophy* 97.1: 178–193

Bourne, C. and Caddick Bourne, E. 2018. "Personification Without Impossible Content". *British Journal of Aesthetics* 58.2: 165–179

Braun, D. 2005. "Empty Names, Fictional Names, Mythical Names". *Noûs* 39.4: 596–631

Chiang, T. 2002. "Division by Zero" in Chiang, T. *Stories of your Life and Others*. New York: Tor Books: 71–90

Cameron, R.P. 2015. "Improve Your Thought Experiments Overnight with Speculative Fiction!". *Midwest Studies in Philosophy* 39: 29–45

Currie, G. 1986. "Fictional Truth". *Philosophical Studies* 50: 195–212

Currie, G. 1995. "Unreliability Refigured: Narrative in Literature and Film". *Journal of Aesthetics and Art Criticism* 53.1: 19–29

Gendler, T. 2000. "The Puzzle of Imaginative Resistance". *Journal of Philosophy* 97.2: 55–81

Geoffroy, Louis. 1841. *Napoléon Apocryphe*. Paris: Chez Paulin.

Gernsback, H. 1950. *Ralph 124C 41+*, 2nd edition. New York: Frederick Fell.

Hanley, R. 2004. "Truth in Fiction: As Good As It Gets". *Australasian Journal of Philosophy* 82.1: 112–128.

Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.

Kripke, S. 2013. *Reference and Existence*. Oxford: Oxford University Press.

Lewis, D. 1978. "Truth in Fiction". *American Philosophical Quarterly* 15: 37–46.

Lewis, D. 1983. "Postscripts to 'Truth in Fiction'" in Lewis, D. *Philosophical Papers Volume 1*. Oxford: Oxford University Press, pp 276–280.

Lewis, D. 2004. "Letters to Beall and Priest" in in G. Priest, J.C. Beall, and B. Armour-Garb (eds.), *The Law of Non-Contradiction: New Philosophical Essays*. Oxford: Oxford University Press, pp 176–7.

Nolan, D. 2007. "A Consistent Reading of *Sylvan's Box*". *Philosophical Quarterly* 57.229: 667–73.

Nolan, D. 2011. "The Extent of Metaphysical Necessity". *Philosophical Perspectives* 25.1: 313–339.

Nolan, D. 2015. "Personification and Impossible Fictions". *British Journal of Aesthetics* 55.1: 57–69.

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

Proudfoot, D. 2018. "Sylvan's Bottle and Other Problems". *Australasian Journal of Logic* 15.2: 95–122

Routley, R. 1979. "The Semantical Structure of Fictional Discourse". *Poetics* 8: 3–30

Salmon, N. 1998. "Nonexistence". *Noûs* 32: 277–319.

Shoemaker, S. 1980. "Causality and Properties" in van Inwagen, P. (ed) *Time and Cause*: Dordrecht: D. Reidel: 109–135.

Swoyer, C. 1982. "The Nature of Natural Laws". *Australasian Journal of Philosophy* 60.3: 203–223.

Walton, K. 1990. *Mimesis as Make-Believe*. Cambridge MA: Harvard University Press.

Williams, J.R.G., and Woodward, R. forthcoming. "The Cognitive Role of Fictionality". *Philosophy and Phenomenological Research*.

Wildman, N. 2019. "The Possibility of Empty Fictions". *Journal of Aesthetics and Art Criticism* 77.1: 35-42

Woods, J. 1974. *The Logic of Fiction*. The Hague: Mouton.

Woods, J. 2018. *Truth in Fiction: Rethinking its Logic*. Dordrecht: Springer

Xhignesse, M. 2016. "The Trouble With Poetic License". *British Journal of Aesthetics* 56.2: 149–161.