# Fictional truth: In defence of the Reality Principle

[Forthcoming in Maier and Stokke (Eds), *The Language of Fiction* (Oxford University Press)]

#### Abstract

A well-known theory about under which circumstances a statement is true in a fiction is The Reality Principle, which originate in the work of David Lewis:

**(RP)** Where p1... pn  are the primary fictional truths of a fiction F , it is true in F that q iff the following holds: were p1 ... pn the case, q would have been the case  (Walton 1990: 44).

RP has been subjected to a number of counterexamples, up to a point where, in the words of Stacie Friend “it is widely recognized that the Reality Principle […] cannot be a universal inference rule for implied story-truths” (Friend 2017: 33). This chapter argues that the strength of these counterexamples is widely overestimated, and that they do not, on closer scrutiny, constitute reasons for rejecting RP.

**Keywords:** Truth in Fiction; The Reality Principle; Gregory Currie; David Lewis; Kathleen Stock; Kendall Walton.

#### Background

Human language involves the capability to talk and think about things that are not, in a wide sense of the term, “here”. We talk and think not only about what is actually the case, but also about what *could* have been the case (if things were different), what *would* have been the case (if things were different in such and such a way), and what *should* be the case (if things were ideal). These modes of discourse are all species of what we with a term from Charles F. Hockett (1960) can call *displacement*, i.e. the capability to talk about what is not immediately present, which Hockett, among others, has claimed to be a constitutive feature of human language.[[1]](#footnote-2)

Another way in which we talk and think about the non-present is through fictions. In story-telling, the events described are not actual, or at least not necessarily so. Doyle can describe the whereabouts of a detective called Sherlock Holmes without thereby claiming to be talking about someone who has ever existed. The events described in fictions are often not real, and the storyteller does not claim that they are.

The issue that constitutes the background to the discussion in this chapter is whether fictive talk is of a species with these other kinds of talking about what is not here. Should we categorise fictions and talk about fictions as yet another kind of displacement, where broadly speaking the same linguistic mechanisms are in play as in those other modes mentioned above? Or, on the contrary, are fictions and talk about fictions a completely different mode of talking about the non-actual, operating on separate or only distantly related principles?

#### The Reality Principle

Thus stated, the issue is both large and fairly abstract. In this chapter, I will be concerned with a narrower and more direct question, which is nevertheless related to this larger background. One way in which talk about fictions has been thought to be intrinsically related to other kinds of displacement concerns truths that are not explicitly stated by the author, but which we nevertheless take to be true in the fiction. For instance, we know that Sherlock Holmes is a detective because that is stated in the story. However, we also take for granted that he has two nostrils, even though this fact is never explicitly stated by Doyle (Lewis 1978). That Sherlock Holmes is a detective is thus an explicit or primary story truth, whereas his having two nostrils is an implicit, or secondary story truth.

Supplementing what is explicitly stated by the authors with such secondary story truths is needed in order to make sense of fictions. The people and things described in stories are not mere backdrops: we assume that there is more to Sherlock Holmes than what is explicitly stated by the author because there aren’t enough explicit story truths to ‘flesh out’ the plot. For instance: a story in which we only assumed the explicit story truths, it would not be true that the characters have a mental life, unless this was explicitly stated by the author.

The question of how implicit story truths are generated has been subject to extensive debate, in which a *locus classicus* is David Lewis’s “Truth in Fiction” (1978). In this article Lewis proposed, but did not, for reasons to be discussed, endorse what has been called the Reality Principle (RP), as the principle whereby fictional truths are generated. The guiding idea behind RP is that truths in fictions are generated by the mechanisms of counterfactual thinking. Accordingto RPthe implicit story truths are those facts that would obtain if every primary story truth were the case. Here is Walton’s formulation:

**(Generic RP)** Where p1... pn are the primary fictional truths of a fiction F, it is true in F that q iff the following holds: were p1 ... pn the case, q would have been the case. (Walton 1990: 144)

To use one of Lewis’s own examples, it seems intuitively true that Sherlock Holmes lives nearer to Paddington Station than to Waterloo Station since nothing in the stories indicates that he does not, and this is where Baker Street is located in the real world. RP gets this kind of example right because, presumably, if all primary fictional truths of the Sherlock Holmes stories were the case, Baker Street would still be closer to Paddington Station than to Waterloo Station. This fact obtains in the fiction because this is what would counterfactually be the case if Doyle’s story was true.

Since RP essentially reduces fictional truth to counterfactual truth—where the primary or explicit story truths correspond to the antecedent of the conditional—it should be clear how it relates to the bigger issue alluded to in Section 1. If RP is true, then the mechanisms of displacement for fictional thought and talk are really the same as those for counterfactual conditionals. Fictional talk and thought would thus be a species of other modes of thinking and talking about the non-actual. Establishing this would shed considerable light on the nature of fictions.

As noted, Lewis himself was agnostic about RP, since he considered it vulnerable to several intuitive counterexamples. Since Lewis, theorists have been less hesitant to discard RP as insufficient for explaining how implicit story truths in fictions are generated. Woodward, for instance, in an introductory article to truth in fiction, asserts that:

“Unfortunately, RP seems to generate the wrong fictional truths.” ( Woodward 2011: 162)

With this claim, Woodward follows, among others, Walton, who in his discussion of the issue maintains that RP is “seriously inadequate” as a principle for how fictional truth is generated (1990: 144). More recently, Stacie Friend has also joined in the criticism:

It is widely recognized that the Reality Principle […] cannot be a universal inference rule for implied story-truths. (Friend 2017: 33)[[2]](#footnote-3)

Other critics of RP include, amongst others, Currie (1990), Badura and Berto (2019), Phillips (1999), and Proudfoot (2006).

#### hypothesis

The thesis for which I argue in this chapter is that the critics of RP have overstated their case and that a variation of the principle discussed by Lewis stands on much firmer ground than has previously been thought. Thus, there is still hope for the idea that fictional truth is generated by the same mechanisms as counterfactual truth.

The version of RP that I will defend is Generic RP, as stated above. It differs from the version discussed by Lewis in two respects, and also from a view often attributed to Lewis in yet another. Here is the version of RP which Lewis discussed:

**(Lewis RP)** A sentence of the form "In the fiction f, " is non-vacuously true iff some world where f is told as known fact and  is true differs less from our actual world, on balance, than does any world where f is told as known fact and  is not true. It is vacuously true iff there are no possible worlds where f is told as known fact.

First, our own version of RP, unlike Lewis’s, makes no mention of a narrator who “tells the story as known fact”. Lewis’s primary reason for taking on board a narrator seems to have been certain problems that arise when combining RP with the causal theory of reference (Lewis 1978: 39). Since there are more immediate worries concerning RP than the consequences it may bring when combined with the causal theory of reference, I have set aside this particular aspect of Lewis’s theory in the present context.

Second, by stating the theory in terms of closeness of possible worlds, Lewis’s version of RP assumes the Lewis-Stalnaker semantics for counterfactual conditionals. Here, by contrast, I will not rely on any theory about the nature of such conditionals. Instead, I will compare our intuitions concerning what is true in various fictions with our intuitions concerning the truth-value of the corresponding counterfactual conditionals, without assuming any specific semantics for the latter. This chapter concerns the relationship between fictional truth and counterfactual thinking, not the relationship between fictional truth and some specific theory of specific theory of the logic and semantics of counterfactual conditionals. One instance where this makes a difference in relation to Lewis’s original theory are impossible fictions, where the Lewis version of RP runs into trouble. The neutral approach, I argue, does not face similar challenges (see § 7).

Third, what Generic RP refers to in terms of “primary story truths”, Lewis’s simply refers to as “the fiction f”. Here, the view sometimes attributed to Lewis is that primary story truths (what forms the antecedent of the counterfactual conditionals) are only the propositions expressed by the sentences of the fiction in virtue of their conventional meaning and, potentially, the logical consequences of these (cf. for instance Stock 2018: 49). I do not think that this is what Lewis had in mind,[[3]](#footnote-4) though since this not an exegetical text I will not indulge in discussions of the interpretation of Lewis. As explained in §4, a more tenable version of RP maintains that that the primary story truths, or “the story told”, are what is *communicated* by the text.

The objections against RP in the literature typically come in one of two forms. Firstly, there are *under-generation* objections, i.e. objections arising from the fact that something is (allegedly) intuitively true in some fiction but nevertheless not recognised as such by RP. §4 and §6 below discuss such problems. They arise, it is argued, by and large from understanding adherents of RP to be committed to an overly-narrow conception of primary story truths.

Secondly, there are *over-generation* objections, which are discussed in §5, §7 and §8 below. These objections consist of claims that something is untrue in a fiction yet predicted to be true by RP. I suspect that much of the motivating reasons for the widespread skepticism of RP are found in this class of objections. In particular, many theorists resist the consequence of RP that there are truths in fictions which the author herself would not know to be true. In §8 I argue that this consequence is not a liability of RP but, on the contrary, something which any plausible theory about truth in fiction must accommodate.

#### Primary story truths and interpretation

Above, I characterised primary story truths as being what is *explicitly* stated by the story, and secondary story truths as that which is *implicit.* While suitable as an initial characterization, this is, to my mind, not the most accurate way to lay out the land.

Consider a classic case of conversational implicature. If I were to report that my friend Watson has 15 pounds in his pocket, the other participants in the conversation would be warranted to infer that, as far as I know, Watson has no more than 15 pounds in his pocket. As in the case of other implicatures, this inference can be cancelled by simply adding that “he even has 20” or something similar. Likewise, if Doyle wrote that Watson has 15 pounds in his pocket, readers would by default infer that Watson has no more than 15 pounds in his pocket. In this respect, speech and writings which report events in a fiction seem to work in the same way as non-fictions. The problem for RP, as it is sometimes understood, is that the fact that Watson has no more than 15 pounds in his pocket, while not *explicitly* stated by Doyle in this scenario, is still not a secondary story truth along the lines of RP. The inference is not generated through counterfactual thinking, but by standard mechanisms of communication.

Bearing this in mind, in the following discussion of RP I will take the primary truths of a fiction to be not only what is explicitly stated in the text, but instead to be constituted by what is *communicated* by it. The notion of communication I rely on here is that of Paul Grice (1989). On the Gricean view, somewhat simplified, P is communicated by an utterance or piece of text iff it is produced with the intention that the receiver comes to believe that P is true in the fiction, on the basis of recognising that the producer has this intention. The primary story truths will, according to this Gricean version, include not only what is explicitly said, but also what is conversationally implied, since conversational implicatures are generated by what the producer intends her audience to believe on the basis of what is explicitly said, and the receiver’s recognition of this.[[4]](#footnote-5)

Armed with this notion of primary story truth, we are now in a position to address one type of criticism which has been directed at RP. Here is an example from Walton:

Recall the suicide of Mrs. Verloc on her voyage to the Continent in Conrad’s Secret Agent. The newspaper headline, ‘Suicide of a Lady Passenger from a cross-Channel Boat,’ informs the reader of her death. But how can we jump so irresponsibly to the conclusion that she was the victim? We have some additional circumstantial evidence, to be sure. We know that Mrs. Verloc was distraught after having killed her husband, and was afraid of the gallows. Earlier she had contemplated drowning herself in the Thames. Ossipon had abandoned her on the train and stolen her money. But little if any of this additional evidence is needed to establish the fact that fictionally it was Mrs. Verloc who jumped from the ferry. And even this evidence would, in a real case, stand in need of confirmation; there could easily have been another suicidal passenger crossing the Channel the same night. It is doubtful at best that, were a newspaper to carry that headline in those circumstances, it would have been Mrs. Verloc who had jumped (or even that it would have been more likely than not that it was she) […] Yet there is no doubt whatever that, fictionally, the suicide was hers. (Walton 1990: 162)

The objection is that RP, as formulated by Lewis, does not capture the fact that Mrs. Verloc commits suicide in the story, since this is neither *explicitly* said nor generated as a secondary story truth by RP. To see the latter, consider the relevant conditional. If everything that is *explicitly* said in the story were true, would it be the case that a suicide from a channel boat spoken about in the British press the next day was hers? Walton’s point is that it would not—the British press could be speaking about some other woman’s suicide. The workings of counterfactual thinking cannot account for our strong inclination to infer that the newspaper headlines concern Mrs. Verloc.

By including what is conversationally implied by the text in the primary story truths, RP is insulated against this line of criticism. The story certainly *implies* that the headlines speak about Mrs. Verlock’s suicide. As noted by Matravers (2014: 82–83), this example has nothing specifically to do with fiction, but rather with communication in general. Imagine that someone told the same story as a narrative about real events, ending with something along the lines of “On the next day, one could read in the newspaper about the suicide of a British woman from a channel boat”. Of course, one would on the basis of such narration infer that the person concerned was Mrs. Verloc. Why would the narrative report the suicide of a British woman from a channel boat if this would be completely unrelated to the previous events described? Plausibly, this implicature would be generated by the expectation that the speaker abides by the maxim of relevance. We would assume that the person making the report is co-operative, so we would take her to intend us to understand that it is indeed Mrs. Verloc who has committed suicide, rather than someone unrelated to the story. When we take it to be true in the story that Ms. Verloc commits suicide, the same mechanisms are at play. Our version of RP accounts for this.[[5]](#footnote-6)

A similar response can be given to another of Walton’s objections to RP. Walton writes:

A storyteller, in a culture in which it is universally and firmly agreed that the earth is flat and that to venture too far out to sea is to risk falling off, invents a yarn about bold mariners who do sail far out to sea. No mention is made in the story of the shape of the earth or of the danger. That would be unnecessary, the teller thinks, for he and his audience assume the earth in the story to be shaped as they believe it is in reality. All take it to be implied that fictionally there is, somewhere in the vast ocean, a precipice to nothingness. (Walton 1990: 150)

In this scenario, Walton thinks that RP would deliver the result that the earth in the story is (still) round, which would be incorrect. However, since, as Walton says, it is *implied* that the earth is flat, our current version of RP is not vulnerable to this objection. By contrast, if this was not even implied, I see no pressure to accept that the earth is not round in the story.

A third kind of objection can be handled along the same lines. This objection concerns hidden messages or signs planted in the text by the author which establishes some things as true in the fiction. An example given by Stock is from Charlottes Brontë’s *Jane Eyre*, where the text’s description of the protagonist’s particularly red-coloured “womb-like” room has induced critics to infer that it is true in the fiction that losing her mother early in life has had a great psychological effect on her (Stock 2018: 53). This inference is not supported by RP, says the objection.

In this and many similar examples,[[6]](#footnote-7) it is certainly beyond reasonable doubt that what the sign or hidden message is taken to communicate really is true in the fiction. But let’s focus on the general issue of whether it is possible for the author to establish fictional truth in this way. There is no reason for an adherent of RP to deny this. Just as the primary story truths may include what is conversationally implied but not explicitly stated, there seems to be no principled reason to deny that the author can communicate things in more roundabout ways, such as through hints and signs in the relevant sense. Again, a more inclusive notion of primary story truths than the one normally attributed to Lewis salvages RP.

Perhaps someone is inclined to think that taking onboard such a broad notion of primary story truths constitutes a form of cheating. But I don’t see what would motivate the focus on what is explicitly said, in contrast to what is communicated, or why an adherent of RP would want to limit her theory in such a way. That more is communicated than what is explicitly said is a general phenomenon pertaining to language use, and until we are given reasons to believe otherwise, it seems reasonable to think that the mechanisms that are at play in moving from explicit statement to communicated content in non-fictional discourse, are also at play in fictional speech and text.[[7]](#footnote-8)

A related worry might be that this broad notion of primary story truths abolishes the need for RP. Why, after all, should we believe that there are additional truths in the fiction, beyond what the author has communicated? The answer is that this notion of primary story truths still does not cover the examples which Lewis originally offered as motivation for RP, like the fact that Baker Street is closer to Paddington Station than to Waterloo Station in the fiction. This is neither stated, nor implied in the relevant sense, by Doyle’s text.[[8]](#footnote-9) As we shall see in §8, RP also accounts for our intuitions concerning the fictional truth of facts of which the author could not possibly have been aware. So in no way does our expanded notion of primary story truth render RP otiose.

#### Unreliable narration

An interesting case against RP emanates from issues concerning interpretation and unreliable narration. Unlike the objections against RP discussed above, the idea is that RP *over*-generates fictional truths in stories with potentially unreliable narrators. The previously mentioned considerations concerning primary story truths are nevertheless here relevant.

Again, the objection goes back to Lewis himself. It has also been developed by, among others, Currie (1990: 66–67) and Friend (2017: 33). Currie’s version is presented here. Henry James’s gothic novel *The Turn of the Screw* tells the story of a young governess who experiences the house where she is working as haunted. Whether the house actually is haunted, or whether the governess rather suffers from some kind of mental illness, has been debated by critics and literary scholars since the publication of the novel. The problem for RP is that it seems to settle the matter in favour of the interpretation according to which she is delusional, since this is what counterfactually would be the case. Currie writes:

But it seems wrong to make our interpretation of James’s story depend so decisively on what we believe about the existence and non-existence of ghosts. Someone skeptical of ghosts is not thereby excluded from believing that *The Turn of the Screw* is a ghost story. (1990: 67)

I agree with this verdict, and also with the claim that it constitutes a potential challenge for RP. To see why, consider the following conditional, where the antecedent is a stand-in for a complete description of the plot in James’s novel:

(1) If there was a governess who reported that a house where she was working was haunted, etc., then it would be true that there are ghosts.

(1) is presumably false. If someone were to report that events like those in *The Turn of the Screw* are real, they would surely be mistaken or lying. Accordingly, RP seems to deliver the verdict that it is false that there are ghosts in the story. As Currie points out, that seems like something we do not want our theory of truth in fiction to settle for us. This is an instance of (potentially) unreliable narration, and the problem that such fictions constitute for RP.[[9]](#footnote-10)

I take the basic data point here to be that there are different ways to read James’s story. The reading in which the house really is haunted is one where we deem the governess as trustworthy and therefore take her story at face value. The other reading is one where we take her not to be trustworthy, and therefore dismiss her reports about ghosts as delusional. We want our notion of truth in fiction to be neutral concerning this issue. This is easy to accommodate within the framework of RP. The two interpretations, I submit, are two different ways of settling what the primary truths of the story are. If we take the primary truths to be *what the governess tells us*, we accordingly insert what she says into the conditionals by means of which the secondary story truths are settled. On this reading, there are ghosts in the story since we take the governess’s reports of the sighting of dead people at face value. Accordingly, we treat their content as primary story truths. On the other, skeptical reading, we don’t take the *content* of what the governess says at face-value, but instead take the primary story truths to be *that it is reported that* she experienced the events in such and such way. That is, we take the frame story, in which a person is reading aloud the alleged testimony of the governess, as being the primary story truths (see Zucchi, (this volume) and Maier and Semeijn, (this

volume)) for approaches to unreliable narration along these lines) On this reading, there are no ghosts in the story, since a more plausible way of accounting for her reports (presumably) is that she suffers from mental illness.

I am inclined to believe that there is no such thing as the right way of interpreting the story in this regard. On this view, the two manners in which to work out what the primary story truths are, are on an equal footing, and so it really is indeterminate whether there are ghosts in the story or not. Nonetheless, one might believe that there really is such a thing as the correct interpretation—perhaps the one which accords with the author’s intentions. If so, one simply considers James’s intention as having authority concerning what the primary story truths are. Our defence of RP can be neutral concerning this issue.[[10]](#footnote-11)

A similar reply can be given to an objection that Lewis himself raised against RP. In *The Story of the Speckled Band*, it turns out that the murder was committed with the help of a Russell’s viper that climbs a rope. However, as has been pointed out in the literature, Russell’s vipers are not constrictors and are therefore unable to scale ropes. According to RP, the objection goes, we are therefore forced to conclude that Holmes didn’t solve the case. If someone did what the narrator of the story, Watson, is doing in the fiction, namely reporting about a murder case involving a rope climbing Russell’s viper, he would surely be lying or be mistaken. Yet, since Holmes did solve the case, RP yields the wrong consequences in this instance, it is alleged (Lewis 1978: 43). But again, including the narrator’s reliability as one of the primary story truths accommodates this objection. If we take Watson to be reliable, by far the most natural way to read the story, the story worlds are worlds in which Russell’s vipers have capabilities which they lack in the real world.[[11]](#footnote-12) If we (less naturally) take the primary story truths to be that Watson tells us these things, then it is of course not beyond doubt that he is sometimes wrong.

#### Genre considerations

Another case in which RP is said to fail to capture instances of intuitive truth in fiction concerns genre considerations. Again, this objection goes back to Lewis himself. Here is Friend’s formulation, with Lewis’s examples:

One problem is that [RP] treats primary story-truths as the only reason for departures from reality. Genre considerations may provide another reason. If in a fantasy story about knights a dragon appears, we are invited to imagine that the dragon breathes fire unless otherwise indicated, even if a world without fire-breathing creatures is closer to the actual world than one with them would be. (Friend 2017: 33)

He is a similar example by Woodward:

A convention governing the genre of ‘zombie’ stories is that zombies do not run. They kind of stumble along, typically stretching their arms towards their victims. This is never explicitly stated, however: it’s a stereotype that has evolved with the development of the genre to the point where some aficionados refuse to classify stories involving ‘running zombies’ as genuine members of the genre. (There are stories – The Return of the Living Dead being an example – in which zombies do run, but this is self‐consciously anti‐conventional. 28 Days Later is a trickier case.) The problem is that even if we take the body of primary fictional truths that are directly generated by a story, it is probably neither true nor mutually believed that, were they all true, zombies do not run. But it is nonetheless fictionally true that they do not. (Woodward 2011: 162; cf. also Stock 2018: 52)

Again, the objections are primarily levelled against the Lewis-version of RP, but I take it that they are also meant to cover other versions.

For these to be effective counterexamples against Generic RP, two things need to be the case. First, counterfactuals like (2) and (3), where the antecedents should again be taken as shorthand for the kind of antecedents generated by the primary story truths of each respective story, should be intuitively false (or at least not true):

(2) If there had been dragons, they would have breathed fire.

(3) If there had been zombies, they would have been unable to run.

Secondly, it must be the case that it would be true in fictions in which there are zombies and dragons, that the dragons would breathe fire and the zombies would be unable to run, although no stance on the issue was taken by the author. That is, it must be that our judgement about the two kinds of cases diverge. I do not think that they do, hence I do not think that they are good counterexamples to RP. Someone who judges (2) and (3) to be false will not take for granted that a dragon in a story breathes fire, if this is not explicitly stated or hinted at in the story. Someone who, like myself takes (2) and (3) to be true will, on the other hand, believe that the creatures possess the relevant attributes in the stories, even if this is not stated.[[12]](#footnote-13)

Of course, one might think that this merely goes to show that these specific purported counterexamples are not effective, but that there are other instances where genre considerations make fictional truth diverge from RP. I think that there is a shortage of good examples to this effect. More generally, there is the risk of conflating two different issues when discussing these things. It is certainly true that the genre of a piece of fiction might affect what expectations we have of what will happen. We would be much more surprised by a description of a supernatural event when reading, say, an Agatha Christie novel, than when reading a fantasy book. For someone who has not read the book, it is more likely that the latter, but not the former, will contain transgressions of the laws of nature. This, however, is not the same as stating that genre-belonging affects what we take to be true in the story, once all the primary story truths are taken into account (a similar point is made by Stock (2018: 72)). These are, however, predictions about what will happen later on in the story.

One further point concerning this matter must be made. Imagine a community which, for one reason or another, is very strict about genre conventions. In this community, the law dictates that every story must include an explicit indication of its genre, and for each genre there is a list of propositions that are true in stories belonging to this genre. Everyone in our imagined community know these lists by heart. I take it that it would be absurd to deny that genre-conventions play a role in determining fictional truth in such a community. But the intelligibility of this scenario does not show is that RP is misguided. What an adherent of RP should say instead, I submit, is that in this community, genre conventions are part of the primary story truths. They should go into the antecedents, rather than the consequent clause of the conditionals which determine truth in fictions. Above, I expressed some skepticism concerning whether our own community in any way approximates the one described in this scenario. Yet, even if I am wrong about this, so that genre-conventions do in fact play a role in determining truth in fictions, this by no means shows that RP is fundamentally misguided. What it would show is that we must sometimes include genre-conventions among the primary story truths.

#### Impossible fictions

Impossible fictions constitute a case where Lewis’s version of RP over-generates fictional truth. Fictions sometimes contain elements which are metaphysically impossible. Gregor’s transformation into a beetle in Kafka’s *Metamorphosis* is sometimes mentioned (Woodward 2011).[[13]](#footnote-14) The problem is that according to the Lewis-Stalnaker semantics for counterfactual conditionals, such conditionals with necessarily false antecedents are vacuously true. This is because there will be no *possible* world where the antecedent is true. The upshot is that impossible fictions combined with Lewis RP make it the case that *anything* is an implicit story truth in such stories. That would be implausible.

The natural rejoinder, mentioned in passing by Walton (1990: 146), is to hold onto RP while rejecting the Lewis-Stalnaker semantics for counterfactuals. As is well known, there are independent reasons for doing so. There are many examples of counterfactuals with impossible antecedents which nevertheless strike us as false. Here is an example from Nolan (1997):

1. If Hobbes had (secretly) squared the circle, all sick children in the mountains of South America at the time would have cared.

Examples like this are widely taken to show that counterfactual conditionals with necessarily false antecedents can be false (and non-vacuously true), and that this therefore offers a reason to reject the Lewis-Stalnaker semantics of such conditionals. By the same token, not everything needs to be true in a fiction with metaphysically impossible storylines. In short, since counterfactual conditionals with necessarily false antecedents can be false and non-vacuously true, there is no case against Generic RP on the basis of impossible fictions as such.[[14]](#footnote-15)

#### Beyond the author

Another alleged problem for RP is that it will import truths from the actual world, of which the author (and potentially the readers) are unaware. Here is Lewis’s illustration, in which we are to suppose that there are, unexpectedly, gnomes in the actual world:

“it seems clear that whatever purple gnomes may be hidden in odd corners of our actual world, there are still none of them in the worlds of Sherlock Holmes. (Lewis 1978: 44)

Woodward elaborates on the same point:

One problem is that a huge variety of actually true propositions are consistent with any story’s primary fictional truths. But if the primary truths don’t rule out p, then RP will entail that p is fictional if it’s actually the case that p. For instance, RP entails that it is true in [Nabokov’s] *Pnin* that there are exactly two trillion stars in the night sky (if there are), simply because nothing the story says is inconsistent with that fact. Similarly, RP entails that it is true in *Pnin* that every mathematical truth holds, no matter how complicated, simply because none of its primary fictional truths are inconsistent with the mathematical facts. Finally, RP entails that it is true in *Pnin* that Napoleon had exactly one million hairs on his head (if he did), simply because nothing the story says is inconsistent with this fact. These results are highly counterintuitive: they imply that not even the story’s author could’ve known what is fictionally true in his work. (Woodward 2011: 162; cf also Currie 1990: 65 and Stock 2018: 51)

To illustrate, RP has these consequences since, presumably, if every primary story truth in *Pnin,* a novel set in the 20th century, were true, Napoleon’s hair would still have been the same.

It is not clear that this is problematic in the way that these theorists suggest. Many people, including myself, are perfectly fine with this result, as far as the intuitiveness of the relevant judgements is concerned. Why, after all, would Napoleon be any different in the world(s) of *Pnin*, than in the actual world, given that nothing is said about his characteristics? In the quote above, Woodward indicates that the underlying issue is the allegedly problematic consequence that the world(s) of the fiction are not completely transparent to the author. The reductio of RP is supposed to be the consequence “that not even the story’s author could’ve known what is fictionally true in his work”.

It seems to me that whether you are an adherent to RP or not, there is heavy pressure to accept that some truths unknown to the author are true in the fiction. Consider the following case: Dostoevsky did not know that consumption is caused by the bacteria *Mycobacterium tuberculosis,* since that was only discovered after the novel was written. Nonetheless it seems very natural to think that this is indeed the underlying cause of Katerina Ivanovna’s disease in *Crime and Punishment*. If you doubt it, consider whether she could have been treated with modern medicine. I take it that everyone agrees that she could have. This implies that we take it that the tuberculosis from which she suffers is the same as the one found in our actual world, and this, in turn, means that her illness in the fiction has plenty of properties which were unknown to Dostoevsky, such as its aetiology. Examples with a similar structure are easy to construct. Such examples show that the sort of transparency for the author, to which Woodward and others allude when rejecting RP, is untenable. Once this is recognised, the problem for RP concerning such cases diminishes.

Another alleged consequence of RP in this vicinity concerns truths taking place outside the time frame of the story. This is one of the most recurring criticisms of RP. Here is Proudfoot’s version of it:

[RP] makes it true in the Sherlock Holmes stories, Greek comedy and any other fictional work that Anglo-American philosophy takes a scientific turn in the late 20th century, since some f-world where Anglo-American philosophy takes a scientific turn in the late 20th century will be closer to the actual world, on balance, than any f-world where it is false that Anglo-American philosophy takes a scientific turn in the late 20th century. This is highly counterintuitive. (Proudfoot 2006: 17; Cf. also Stock 2018: 55)[[15]](#footnote-16)

Similarly, Walton maintains that, according to RP, “it will be fictional in ‘Goldilocks and the Three Bears’ that Tensing and Hillary achieved the first ascent of Mount Everest” (1990: 48).

As it stands, this criticism misses its target. It is simply not true that RP, even on the Lewis-Stalnaker semantics, imports all actual truths located outside of the timeframe of the story, into the fiction. To see why, it is instructive to consider an objection which was raised against the Lewis-Stalnaker semantics of counterfactual conditionals voiced by Fine and others. In Fine’s version, it goes as follows:

The counterfactual "If Nixon had pressed the button there would have been a nuclear holocaust" is true or can be imagined to be so. Now suppose that there never will be a nuclear holocaust. Then that counterfactual is, on Lewis's analysis, very likely false. For given any world in which antecedent and consequent are both true it will be easy to imagine a closer world in which the antecedent is true and the consequent false. For we need only imagine a change that prevents the holocaust but that does not require such a great divergence from reality. (Fine 1975: 452)

The kind of world which Fine has in mind is a world where Nixon pushes the button, but, by an intervention of a miracle, nothing happens. Lewis responded to this objection by pointing out that the relevant notion of “similarity” (and therefore also that of closeness of worlds) is elusive, and that the informative things that we can say about it will be based on our intuitions regarding the truth-values we ascribe to various counterfactual conditionals, like the one in Fine’s example. In essence, Lewis’s diagnosis was that the similarity relation we make use of when thinking about Fine’s example, and perhaps most counterfactuals, is one which prioritises maximising similarity of particular facts in relation to our actual world only up to the time shortly prior to the event described in the antecedent. The closest worlds are worlds which are like this, but which after the event(s) of the antecedent continue to develop on the basis of *laws* that are as similar to our own world as possible, rather than along with *facts* that are similar to our actual ones. We take Fine’s conditional to be true since the closest worlds where the antecedent is true, are worlds in which the past prior to the antecedent is similar to ours, and which is governed by similar natural laws as our own, and in which there is a nuclear holocaust. This is so even if such worlds differ substantially from our own concerning particular matters of fact *after* the event described in the antecedent (because of the instigated nuclear holocaust) (Lewis, 1979).

In a similar fashion, when processing what would be the case if a particular fiction were true, we do not seek to maximise similarity to our own world concerning particular matters of fact after the event of the antecedent. To see this, just consider the intuitive truth-value of counterfactual conditionals which describe some of the primary story-truths of Greek comedies and Goldilocks respectively:

1. If there had been gods and monsters interfering with people’s everyday lives, Anglo-American philosophy would still have developed in the same way.
2. If there had been talking bears, Anglo-American philosophy would still have developed in the same way.

Neither (5) or (6) is, I take it, intuitively true. Worlds in which these antecedents are true would be substantially different from our own, and for this reason we do not find it intuitive that such worlds would accord much with our own concerning particular matters of fact in their corresponding future, as, for instance, facts concerning dominating trends in Anglo-American philosophy. Since (5) and (6) are intuitively not true, it is not the case that adherents of RP are committed to the truths of their consequent clauses in the relevant stories, as claimed by Walton, Proudfoot and others.

There are more moderate variations of the same objection. Badura and Berto (2019) invite us to consider the statement that, in the Sherlock Holmes stories, Trump wins the US presidential election in 2016. They write:

By [RP], this is true iff some [Sherlock Holmes] world w in which Trump wins the election in 2016 is closer to our world than is any [Sherlock Holmes] world where Trump does not. With there being such a w, one is saying that [Sherlock Holmes] stories are, and were at the time of writing, about Trump and our future. But how can the [Sherlock Holmes] stories be about (say) flying cars in 2160? Call this the aboutness-objection. (Badura and Berto 2019: 181)

In this version of the objection, it is not claimed that *all* stories import future truths from the actual world, but only that this is the case with some stories, like Sherlock Holmes. Perhaps they think that stories that are “realistic”, in the sense that accommodating their primary truths does not demand a widespread change in the laws governing the world (in comparison to our own), will have futures which to a large extent resemble our own. They will therefore have the allegedly problematic consequence that the actual future will to a large extent be imported into them. There are several things to say about this. First, for our present concerns, this objection only hits its mark if it is true that:

1. If all the primary story truths in the Sherlock Holmes stories were true, Trump would still have won the 2016 US presidential election.

It seems reasonable to be somewhat agnostic about the truth-value of (7). We simply do not know how the Sherlock Holmes stories being true would have affected the future course of events. But let’s assume that (7) is indeed true, and that, for this reason, it really is true in the fiction that Trump wins the election in 2016, according to RP. Would that be a *reductio* of the principle? Surely not. On the supposition that the primary story truths of the Sherlock Holmes stories are indeed compatible with our actual future, there seems to be no reason to deny that the future of the Sherlock Holmes stories is indeed our actual future. Why, on this supposition, would it be any different? At the core of Badura and Berto’s objection above seems to lie the worry that this would make the Sherlock Holmes stories be *about* Trump’s election win. There is, however, no pressure for an adherent of RP to accept that. Not everything assumed to be true by a narrative, fictional or non-fictional, is what that narrative is about. It is true in the Sherlock Holmes stories that he plays the violin using a bow, but that is certainly not what these stories are about. Similarly, the Sherlock Holmes stories are *about* a detective, not about Trump, even if future truths are imported into the fiction from the actual world.

I conclude that its prominent position in the literature notwithstanding, the future-truths objection against RP fails.

#### Conclusion

These are, I take it, the most prominent objections to RP. It should perhaps be noted that I have not provided much positive motivation for RP, or argued for why it is preferable to other theories about truth in fiction. An extensive discussion regarding this will have to wait for another occasion, but to say something about it in the present context I think examples like the one involving the aetiology of consumption in Dostoevsky’s novels, briefly discussed in §8, are such that one needs RP, or at least something in the close vicinity, to account for them. In the meantime, I hope that the present discussion has shown that the strength of the objections to RP is overestimated, and that the general idea which underlies the principle, that fictional truth is generated by the mechanisms of counterfactual thinking, is therefore still very much a live option.

# references

Badura, C., Berto, F., 2019. Truth in Fiction, Impossible Worlds, and Belief Revision. Australas. J. Philos.

Currie, G., 1990. The Nature of Fiction. Cambridge University Press.

Davies, D. (1996). Fictional Truth and Fictional Authors. *British Journal of Aesthetics*, *36*(1), 43–55.

Fine, K., 1975. Critical Notice of Lewis, Counterfactuals. Mind 84, 451–458.

Friend, S., 2017. The Real Foundation of Fictional Worlds. Australas. J. Philos. 95, 29–42.

Grice, P., 1989. Studies in the way of words, 3. print. ed. Harvard Univ. Press, Cambridge, Mass.

Hanley, R., 2004. As Good As It Gets: Lewis on Truth in Fiction. Australas. J. Philos. 82, 112–128.

Hockett, C.F., 1960. The Origin of Speech. Sci. Am. 203, 88–96.

Lewis, D., 1983. Postscript to Truth in Fiction, in: Philosophical Papers. Oxford University Press, pp. 276–280.

Lewis, D. 1978. Truth in Fiction. Am. Philos. Q. 15, 37–46.

Lewis, D. 1979. Counterfactual Dependence and Time’s Arrow. Noûs 13, 455–476.

Maier, E., Semeijn. M. This volume. Extracting Fictional Truth form Unreliable Sources.

Matravers, D., 2014. Fiction and Narrative. Oxford University Press.

Nolan, D., 1997. Impossible Worlds: A Modest Approach. Notre Dame J. Form. Log. 38, 535–572.

Øhrstrøm, P., Hasle, P., 2015. Future Contingents, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Phillips, J.F., 1999. Truth and Inference in Fiction. Philos. Stud. Int. J. Philos. Anal. Tradit. 94, 273–293.

Proudfoot, D., 2006. Possible Worlds Semantics and Fiction. J. Philos. Log. 35, 9–40.

Sainsbury, M. [Fictional worlds and fiction operators](https://www.marksainsbury.net/uploads/9/7/2/9/97295048/martich-11_sainsbury.pdf), 2014. , in: Empty Representations. Manuel Garcia-Carpintero and Genoveva Marti (eds.) Oxford University Press.

Stock, K., 2018. Only Imagine: Fiction, Interpretation and Imagination. Oxford: Oxford University Press.

Walton, K.L., 1990. Mimesis as Make-Believe: On the Foundations of the Representational Arts. Harvard University Press.

Williamson, T., 2007. The Philosophy of Philosophy. Wiley-Blackwell.

Woodward, R., 2011. Truth in Fiction. Philos. Compass 6, 158–167.

Zucchi, S. This volume. On the generation of Content.

1. Hockett used the term primarily to refer to spatial and temporal displacement. [↑](#footnote-ref-2)
2. In her (2017), Stacie Friend defends a cousin of RP, the *Reality Assumption*: “the assumption that everything that is (really) true is also fictionally the case, unless excluded by the work.” I do not think there is much disagreement between the version of RP I defend and her and Friend’s view. Where we do differ is that Friend doubts that there could be a plausible and finitely stateable principle concerning how truth in fiction is generated (Friend 2017: 34). While perhaps coarse-grained, the version of RP defended here is certainly finitely stateable. [↑](#footnote-ref-3)
3. Lewis writes that “so far as I know it is never stated or implied in the stories themselves that Holmes lives nearer to Paddington” (Lewis 1978: 41). This indicates that he thought both of what is stated and what is implied as being part of the “story being told”, i.e. as primary story truths, as opposed to the secondary story truths which are neither stated nor implied. [↑](#footnote-ref-4)
4. This approach raises an important issue concerning *who* does the implicating. Is it the actual author of the story, or does every story have a fictional narrator who does the narrating, as argued by for instance Currie (1990)? For the purpose of this chapter, I sidestep this issue and proceed as if the author is the person whose reflective intentions are relevant for settling what the primary story truths are. Nothing much hangs on this. A similar account of what the primary story truths are can be generated if the narrator is instead assumed to be a fictional. [↑](#footnote-ref-5)
5. Hanley (2004) defends a similar version of RP. Davies (1996: 44) also thinks that that problem of truth in fiction concerns what, *in addition* to what is stated *and implied* by the text, is true in a fiction. [↑](#footnote-ref-6)
6. Further examples can be found in Stock (2018: 53–54) and Walton (1990: 165). [↑](#footnote-ref-7)
7. Sainsbury seems to interpret Lewis as holding the view concerning primary story truths that I defend in this section. He recognises that this means that interpretation will be needed to establish what the primary story truths are, and he seems to think that this is a problem for this view (Sainsbury 2014: 283-284). Nonetheless, I do not see what the problem is supposed to be. Potentially, Sainsbury thinks that this abolishes the need for RP since no further fictional truths are needed, beyond those retrieved from the process of establishing what the author has communicated. For the reasons given in the main text, this is not correct. [↑](#footnote-ref-8)
8. Again, Lewis himself recognised as much. See footnote 2. [↑](#footnote-ref-9)
9. It is perhaps relevant to mention that in *The Turn of the Screw*, the story is told by a man in a frame story who claims to be reading from a testimony of the events, written by the governess. [↑](#footnote-ref-10)
10. Stock claims that Lewis’s view is “thoroughly anti-intentionalist” and maintains that it cannot be made compatible with intentionalism (2018: 52). As our discussion has shown, this is only true if with “intentionalism” one has in mind the view that what is true in a fiction is *exhausted* by the author’s intentions. RP can certainly accommodate that the author’s intentions are part of determining fictional truth, as in the case of implicatures in the text. [↑](#footnote-ref-11)
11. Lewis writes that “The story never quite says that Holmes was right that the snake climbed the robe (1978 : 43)”. That is not accurate. Right at the end of the story, Watson says that “Such are the true facts of the death of Dr. Grimesby Roylott of Stoke Moran”, thereby affirming that Holmes’ theory about the snake is the right one. Of course, even if he had not said this, it would nevertheless be heavily implied by the story that Holmes is right. [↑](#footnote-ref-12)
12. Lewis writes that “It just might be analytic that nothing is a dragon unless it breathes fire. But suppose I never called Scrulch a dragon; I merely endowed him with all the standard dragonly attributes except fire-breathing” (1978: 45). I take it that this would be a situation in which it is at least implied that Scrulch is a dragon. [↑](#footnote-ref-13)
13. This is meant to be a separate problem to that of inconsistent fictions, discussed in (Lewis 1983). [↑](#footnote-ref-14)
14. One might wonder how the theory would stand if “vacuist” hard-liners, who think that (4) and other similar examples are, contrary to appearance, vacuously true, are right (for instance Williamson 2007: 175). On the one hand, RP would have the problematic consequences described, but on the other hand, on this supposition, we would be systematically misguided in our judgement concerning the truth-value of counterfactual conditionals. On RP, it would then be as expected that we are also systematically misguided concerning truth in fiction (in, for instance, not taking everything to be the case in *Metamorphosis*). So it is unclear whether there really is a problem for RP here, even on such a supposition. [↑](#footnote-ref-15)
15. One can certainly have misgivings about whether it’s true in the actual world that Anglo-American philosophy took a scientific turn at the end of the 20th century. I’m putting such worries to the side. [↑](#footnote-ref-16)