Implicating fictional truth

Nils Franzén¹

Abstract

Some things that we take to be the case in a fictional work are never made explicit by the work itself. For instance, we assume that Sherlock Holmes does not have a third nostril, that he wears underpants and that he has never solved a case with a purple gnome, even though neither of these things is ever mentioned in the narration. This article argues that examples like these can be accounted for through the same content-enriching reasoning that we employ when confronted with non-fictional discourse, with the important difference that fictional discourse essentially involves pretence. Fictional discourse works in much the same way as non-fictional discourse, and what is conveyed without being stated can accordingly be explained through familiar pragmatic mechanisms. It is argued that this account carries some distinct advantages over competing views.

Keywords David lewis · Fictionality · Kathleen stock · Stacie friend · The reality principle · Truth in fiction

1 Introduction

When engaging with fictional works, we have to make inferences concerning what's the case in the fiction that goes beyond what we are explicitly told by the narrator. For instance, we are explicitly told that Sherlock Holmes is a detective and that he lives in an apartment on Baker Street together with his friend Watson. But there are many other things which are not mentioned that we nevertheless assume to be the case. We assume that the London in which many of the stories are set is a city inhabited by millions of people, going about their lives at the same time as Sherlock and Watson

✉ Nils Franzén
   nils.franzen@filosofi.uu.se

1   Department of Historical, Philosophical and Religious Studies, Umeå University, Umeå,
   Sweden

are solving their cases. Likewise, we assume that Sherlock and Watson have internal organs and blood circulation, even though this is not explicitly stated by the narrator. If we did not, the stories would be set in an uncanny parallel universe to ours, where London is desolate and people are not of flesh and blood, but mere empty shells. To make sense of narratives, we must go beyond what is explicit. We must flesh out the story in such a way that things and characters and events in the fiction do not appear as mere theatrical backdrops.

The topic of this article is how we go about this "fleshing out". How do we move from the relatively brief explicit descriptions that we are provided with by the narrator to what is implicitly the case in the fiction? What principles are we operating by when we are making inferences to arrive at implicit "truth in fiction"?

Several different views have been put on the table regarding this. Lewis (1978) famously proposed a counterfactual analysis,[1] which, following Walton (1990) and Friend (2017), I will call the Reality Principle. Here is Walton's rendering of it, where "Primary fictional truths" roughly corresponds to what is reported by the narrator (see § 7 for further discussion):

> **The Reality Principle**: Where $p_1 \ldots p_n$ are the primary fictional truths of a fiction F, it is true in F that q iff the following holds: were $p_1 \ldots p_n$ the case, q would have been the case. (Walton, 1990, p. 144)

For instance, the Reality Principle predicts that it will be a secondary fictional truth that Holme's home on Baker Street will be closer to Paddington station than to Waterloo station, since that *would be* the case if there were a detective named Sherlock Holmes living on Baker Street.

There are other views. For instance, other theorists have suggested that these inferences are generated by what we (the readers) think the author intends us to imagine (Byrne, 1993; Stock, 2017). Making inferences about implicit truth in fiction is to, in broadly speaking Gricean fashion, try to figure out what the author implicitly wants us to imagine is the case on the basis of what is explicitly said.

My purpose here is to suggest a novel answer to the question of implicit truth in fiction and argue that it has some distinct advantages compared to competing accounts. The guiding idea is that fictional discourse is in fact more similar to non-fictional discourse than what has normally been assumed in the literature. Accordingly, a wide of examples implicit truth in fiction can actually be accounted through conversational pragmatics. The basic elements of the view are this: the inferences we make about implicit truth in fictions are, with some exceptions (see § 6), generated by conversational implicatures and similar pragmatic mechanisms. These implicatures are elicited under the mode of pretence since the author is not really asserting, but only pretending to assert (and we, the addressees, are also pretending that she is doing so).

Neither of these elements is new. That (some) implicit truth in fiction is generated by conversational implicatures has been suggested by, e.g., Franzén (2021) and Matravers (2014). That fictional discourse is essentially pretence has been suggested

---

[1] Lewis himself ultimately did not fully endorse the counterfactual analysis, but instead thought that when working out the background truths of fiction, we oscillate between this principle and others.

by, among others, (Searle, 1975; Lewis, 1978). What is new in the current approach is the combination of these two elements, and, moreover, the attempt to explain a wide range of cases of implicit truth in fiction as conversational implicatures and other pragmatic inferences. In particular, the appeal the neo-Gricean notion of generalized conversational implicatures and "stereotypicality inferences" plays an important role in the account.

Before getting started, let me make a proviso. In the following, I move rather freely between talking about, on the one hand, truth in fiction and, on the other hand, the inferences that we make to the effect that something is true in the fiction. This is a simplification. The two notions may come apart, not only because we are sometimes careless or inattentive but also because the author intentionally makes us believe that something is the case in a fiction while in fact it is not. An often-mentioned example of this is Agatha Christie's *The Murder of Roger Ackroyd*, where the narrator is deliberatively misleading the reader and, in the final pages, actually turns out to be the culprit. Since there might always in principle be possible that we are being misled by the author/narrator, any direct move from what we are being led to believe, the inferences that we make about what is true in the fiction, to what is actually true, is illegitimate. In the following, I will bracket this concern and focus on the issue as formulated above, namely that of identifying which principles and processes underlie our inferences about what is implicitly true in fictions. I return to the issue of how to formulate a definition of truth in fiction on this basis in § 7.

2 Pretended pragmatics

To get us started, let me begin with an example of implicit truth in fiction from the literature where I think that it's obvious that what is going on is a case of conversational implicature. It's a case from Walton's (1990), which he uses to argue against Lewis's Reality Principle, stated above:

> 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, p. 162)

The idea behind the objection is that it is never stated in the novel that the suicide is Mrs Verloc's, but it is nevertheless an implicit fictional truth that this is the case. The Reality Principle does not predict this, since, even if everything explicitly described was the case, it is not certain that the suicide described in the newspaper headline would have been hers.

As pointed out by Matravers (2014; see also Franzén, 2021) the implicit story truth that the suicide was Mrs. Verloc's is very naturally accounted for as conversational implicature. Briefly, through the supposition that the narrator abides by the maxim of relation ("Be relevant"), and the mutual recognition of this expectation, the narrator has conversationally implied that they are not describing another suicide, unrelated to the narrative, and that it is therefore Mrs. Verloc's. To strengthen the case that implicit story truths like these are implicatures, one could consider what inferences we would draw if we read the story not as a fictional text, but rather as a report about real events. We would of course draw the conclusion that the narrator intended us to believe that the suicide was Mrs. Verloc's, based on the exact same reasoning. It would be strange if these two inferences (the fictional and the non-fictional) were based on completely different mechanisms, given their similarity.

Once the notion of there being pragmatic implicatures in fictions is on one's radar, it is easy to see that they come in abundance. Just consider a simple scalar implicature. If, in the Sherlock Holmes stories it is reported that Holmes has 2 pounds in his pocket, that seems to be implicating that Holmes has *only* 2 pounds in the same way as if he asserted this about a friend of his in non-fictional circumstances. The underlying idea to be presented here is grounded in this basic observation, that fictional discourse seems very much continuous with non-fictional discourse with respect to pragmatics.[2] Just as the narrator of *The Secret Agent* would have implicated that the suicide was Mrs. Verloc's in a corresponding non-fictional scenario, the narrator of the Sherlock Holmes stories (Watson) would be implicating "only 2 pounds" in the exact same way in the fictional and non-fictional scenarios.[3]

Moreover, looking at the matter from a functional perspective, it should be clear that the explanatory grounds for the phenomenon, the reasons for why conversational implicatures exist, are very much present in the case of fictional discourse as well. Levinson suggests that implicatures arise from the fact that phonetic articulation is a bottleneck: it takes more time to speak than it does to process information. As Levinson puts it "inference is cheap, articulation expensive". Implicature arises out of the human need for a system of communication that maximizes inference at the expense of explicit codification. (Levinson, 2000, p. 29). This reasoning, of course, applies to fictional discourse just as much as it does to assertion. It would be extremely inef-

---

[2] This argument for that fictional discourse is a mode of pretence is similar in spirit to Predelli's (2020) "Uniformity argument" for the pretence theory.

[3] One might wonder if it is Conan Doyle or the narrator of the story, Watson, who is doing the implicating? On the Lewisian view I adopt below, it's Conan Doyle, pretending to be Watson, who is pretending to tell us about the story, and therefore also pretending to be implicating. Watson, of course is really asserting and really implicating (in the fictional story). In the case of so-called "heterodiegetic narration", where there is no narrating character in the story, the author is simply pretending to know about these things herself, alternatively, pretending to be some unknown person knowing about the events.

fective to articulate fiction where all information is explicitly codified and nothing beyond the explicit is even implied.

As stated above, the primary goal here is to argue that a wide array of cases, previously raised in support of other theories about truth in fiction, can in effect be accounted for as conversational implicatures and similar pragmatic mechanisms. But before getting to that part of the argument, a principled question should be addressed.

It is widely held that the author of a fictional text is not *asserting*. In writing Sherlock Holmes, Conan Doyle is not reporting about what the world is like and he is not making claims that can be true or false. The theory of conversational implicatures, by contrast, was developed specifically for assertive discourse, and Grice himself held that the maxims of conversation would have to be partially different for non-assertive discourse (e.g., Grice, 1989, p. 28). This seems to spell trouble for the idea that implicit truth in fiction can be established by the author conversationally implicating in the course of writing her story.

There are several ways of addressing this issue, corresponding to different ideas about what kind of speech act the author is performing in the course of writing fiction. Following our guiding idea of fictional discourse being structurally similar to the assertion in terms of pragmatics, here is the one I want to go for: the author is *pretending* to perform a speech act. So, for instance, Conan Doyle is pretending to be Watson, recounting for us the events of the Sherlock Holmes stories as if they had really happened. This idea has been around in the philosophical literature on fiction and fictional discourse for quite some time (see, e.g., Kripke, 2013; Lewis, 1978; McDonald, 1954; Recanati, 2018; Searle, 1975). However, it has not been put to the kind of use to which I want to put it here, which is to explain how fictional discourse can carry pragmatic content in much the same way as assertions: just as we identify the explicit story truths with what the author is pretending to explicitly tell us by her pretend assertions, we identify implicit story truths with the (pretend) pragmatic content that the author is pretending to conversationally implicate by her pretend assertions.

The pretence-idea is helpfully developed in the way of Walton (1990), who takes fiction to be "games of make-belief" that involve both speakers and addressees. Here is how Lewis, following this idea of Walton's, puts it:

> […] the storyteller's pretence of truth and knowledge is only the tip of the iceberg. There is a cooperative game of make-believe, governed by conventional understandings, with players in (at least) two roles. The storytellers pretend to pass on historical information to their audience; the audience pretends to learn from their words, and to respond accordingly. (Lewis, 1983, s. 291)

The idea that literary fictions are constituted by games of cooperative pretence allows us to explain how the author can pragmatically implicate just as if she were making assertions. The author is pretending to assert (or, sometimes, to perform some other speech act) and the reader pretends to pick up on those assertions, and accordingly also pretends to pick up on the implications of those pretended assertions. To put it another way: insofar as one conceives of literary fictions as games of pretence in this way, it is to be expected that fictional discourse carries pragmatic implicatures

in the same way as assertive discourse, since implicating is a central feature of such discourse.

It should be recognized that the theory that fictional discourse is essentially pretence is not uncontroversial and has been subjected to criticism (Bergman & Franzén, 2022; Currie, 1990). But this criticism is directed at the notion that the author's discourse is *sufficiently* characterized as a pretence, since it fails to distinguish it from other modes of fictive discourse, like the actor's discourse on stage. This criticism is compatible with pretence being a necessary feature of the author's discourse.

On the basis of this, let's call the ensuing view about truth in fiction the "pretended pragmatics view", or "PP-view" for short. While I will end up endorsing a more qualified view (§ 7) we can tentatively think of the view as making a universal claim: all implicit truths in fictions are in fact explainable through (pretended) pragmatics.

3 Stereotypicality inferences

In this section, I will argue that an array of examples that Lewis invoked in his original discussion of truth in fiction can be accounted for by the PP-view. Contrary to what Lewis thought, these examples do not support fictional truth going beyond what is stated and implied by the author. Let's look at what Lewis said:

> For example, I claim that in the Holmes stories, Holmes lives nearer to Paddington Station than to Waterloo Station. A glance at the map will show you that his address in Baker Street is much nearer to Paddington. Yet the map is not part of the stories; and so far as I know *it is never stated or implied* in the stories themselves that Holmes lives nearer to Paddington (Lewis, 1978, p.41, my italics).

If we interpret "implied" in the quotation above as "implicated", rather than "entailed", it seems like Lewis already was on board with the notion that truth in fiction can be established through implicatures. Be that as it may, he thought that we nevertheless need the Reality Principle (see § 1) to account for, for instance, the implicit truth that Baker Street is nearer to Paddington than to Waterloo in the novel. Here are some other examples of story truths that are neither "stated [n]or implied" according to Lewis:

- Sherlock Holmes does not have a third nostril.
- He wears underpants.
- He has never had a case in which the murderer turned out to be a purple gnome.
- He solved all his cases without the aid of divine revelation.
- He has never visited the moons of Saturn.
- A dragon in a fiction can breathe fire, even if this is never mentioned explicitly.

As a point of departure for our own pretended pragmatics account of these examples, we will use Levinson's (2000) neo-Gricean theory of conversational implicatures. Levinson postulates a "third layer" of meaning, in-between speaker meaning – what

the speaker intends to communicate – and the conventional meaning of an utterance. This layer of "utterance type meaning" is, in Levinson's words, "[…]a level of systematic pragmatic inference based *not* on direct computations about speaker-intentions but rather on general expectations about how language is normally used" (Levinson, 2000, s. 22, Levinson's italics). Among other aspects of meaning belonging to this level are what Levinson, following Grice, calls "generalized conversational implicatures". These are conversational implicatures that are not generated by specific features of the context of utterance, but which belong to the utterance by default, in the absence of defeating circumstances. Generalized conversational implicatures are, in Levinson's theory, derived from the following three "heuristics" (based on Grice's original four maxims):

The Q-heuristic: What isn't said, isn't.

The I-heuristic: What is simply described is stereotypically exemplified.

The M-heuristic: What is said in an abnormal way isn't normal.

We need not go into the specifics of Levinson's theory, but here is an example. A standard scalar implicature is, in Levinson's framework, derived from the Q-heuristic:

(1) Sherlock Holmes has 2 pounds in his pocket.

The implicature that Holmes has *only* two pounds in his pocket is induced by the fact that the speaker did not use a contrasting, stronger expression, e.g., "3 pounds". From the fact that a stronger expression was not used and the Q-heuristic, it can be inferred that Holmes has only 2 pounds in his pocket.

What is important for our purpose is the I-heuristic and the stereotypical inferences it gives rise to. Among the implicatures that the I-principle triggers are that individual objects, persons, and situations are "stereotypically exemplified" if we are not told otherwise. Levinson's own example (originating in the early 80's) is that:

(2) He smiled at the secretary.

implicates that he smiled at the *female* secretary since secretaries are typically women. It should be clear how this is relevant to the examples that Lewis took to motivate the Reality Principle. Simply by omitting to mention that the secretary is in possession of a salient non-stereotypical trait, the speaker has implicated that this is not so. Analogously, Watson has implicated that Sherlock Holmes does not have a third nostril and that he wears underpants, simply by not mentioning this while describing Holmes's looks and habits. In a similar fashion, by omitting to describe any peculiarities of the London in which much of the Sherlock Holmes stories are set, we assume that it instantiates much the same properties as in the real world, like its salient geographical features. So, we assume that Holmes's home at Baker Street is closer to Paddington station than to Waterloo station in the novel.

Next, consider another type of implicit story-truth that has been discussed in the literature, which Lewis invoked as a *problem* for the Reality Principle, namely of "intra-fictional carry over":

"Suppose I write a story about the dragon Scrulch, a beautiful princess, a bold knight, and what not. It is a perfectly typical instance of its stylized genre, except that I never say that Scrulch breathes fire. Does he nevertheless breathe fire in my story? Perhaps so, because dragons in that sort of story do breathe fire. But the explicit content does not make him breathe fire." (Lewis, 1978, p. 45).

The case is meant to be a problem for the Reality Principle. If a creature with Scrulch's explicitly described characteristics had existed, it would not have breathed fire. Again, this kind of case is readily explained as a stereotypicality inference. While it might not be analytical that dragons breathe fire, a stereo-typical dragon does breathe fire. Since we are not told otherwise, we assume that Scrulch is a stereotypical dragon.[4]

Lewis's other cases, for instance, that he has not solved a case involving divine revelation, are not naturally described as concerning properties of individuals, and are therefore not derivable as I-implicatures. Still, I want to argue, they concern a kind of inference we make on the basis of analogous reasoning. To see this, we can again consider a non-fictional case. Yesterday, during a 30-minutes call to my brother, he did not report that his partner just died unexpectedly, so I, very naturally, assume that his partner has not just died unexpectedly. It's simply one of those things that I would expect him to mention, to say the least. While it is perhaps not right to say that my brother has conversationally implicated that his partner has not died unexpectedly (nothing in the vicinity was discussed), it is a very natural pragmatic inference to make on the basis of what was said and what was not.

Similarly, we would certainly expect Watson, who is dedicated to telling us about the cases of Sherlock Holmes (so we are pretending), to tell us about any incidents involving purple gnomes, visits to Saturn, or divine revelation. Watson is in the business of chronicling the cases of Sherlock Holmes. When doing so, we expect him to not withhold some extremely extraordinary event that has occurred to him. Not telling us about Holmes's recent visit to Saturn, for instance, would be highly misleading.

If what I argued in this section is correct, the PP-view can account for the array of examples that Lewis invoked in his original discussion of truth in fiction. Contrary to what Lewis himself, and many after him have thought, the examples he offered did not motivate implicit truth in fiction going beyond what is stated and pragmatically implied by the author.

Next, let's compare the PP-view to some other proposals about how implicit truth in fiction is generated.

---

[4] Lewis also suggests that even if it's not mentioned that Scrulch is a dragon, as long as the text "endowed him with all the standard dragonly attributes except fire-breathing", it would still be the case that he breathes fire. I take it that in such a case, it is at least implied that Schrulch is a dragon.

4 Other theories

Friend (2017), following Walton (1990, Chap. 4) thinks, that there are multiple of "principles of generation" of truth in fiction and that which ones are suitably employed with respect to a given work is highly context-sensitive. Preceding any such specific interpretational principles is the Reality Assumption, which on Friend's view provides the default starting point of any interpretation:

> **The Reality Assumption**: Everything that is (really) true is also fictionally the case, unless excluded by the work.[5]

The Reality Assumption is motivated by a familiar class of examples, such as that it's true in Jane Austen's *Emma* that human beings have arms and legs, even though this is not explicitly stated, and that people's behavior is generally explained by the standard principles of human psychology if nothing else is indicated (Friend, 2017, p. 32).

The PP-view agree with Friend that we are guided by something like the Reality Assumption in our engagement with fictional works, and that this is so to speak a default inference. Where it goes beyond anything that Friend argues is in not taking the Reality Assumption to be a *sui generis* feature of fictional discourse. Instead, the assumption is treated as deriving from more general features of human communication and the stereotypicality inferences outlined in the previous section. That a PP-view can explain why the Reality Assumption is operative in our engagement with fiction is an important feature of the view.

This difference also applies more generally. Again, Friend, following Walton, thinks that there is an abundance of principles of generation of fictional truth that are not statable in "a single, finitely characterized rule" (Friend, 2017, p. 34). What I have argued here does not disprove that notion, but as with the example from Conrad's *The Secret Agent* in section § 2, many of the examples that Walton offers in favor of this view are actually derivable as conversational implicatures (see Franzén, 2021, for further discussion). As such they do not speak in favour of a multiplicity that cannot ultimately be explained through the pragmatics of discourse. Of course, that does not show that all of Walton's examples for the multiplicity of principles of generation can be accounted for in this way. I will not try to adjudicate the matter here, but just note that this is where the line of demarcation runs between the pragmatic view and the Friend-Walton view.

Another group of theorists take as their point of departure the notion that the fictional speech act is an *invitation to make-believe* (e.g. Byrne, 1993; Currie, 1990; Stock, 2017). What is true in the fiction, on this family of views, is what the author intends the reader to imagine. Make-believe views share with the PP-view a broadly Gricean point of departure. According to make-believe views, the author has a reflex-

---

[5] The formulation is close to what Lamarque and Olsen call the "The Principle of Verisimilitude": "Fictional states of affairs (objects, events, personages) can be assumed to be like ordinary states of affairs (objects, events, personages) failing indications to the contrary" ((Lamarque & Olsen, 1994, p. 95). What I say about the Reality Assumption in the main text applies equally to the Principle of Verisimilitude.

ive intention that her utterance should cause her reader to imagine a specific content and do so by virtue of the reader's recognition that this is what the author intends her to do. Truth in fiction, by this conception, roughly corresponds to Grice's concept of speaker meaning, with the difference that the purpose of the linguistic act of the author is that the reader should make-believe or imagine, rather than believe, what is speaker-meant.[6]

A salient difference between the PP-view and make-believe views is that the latter appeal to a special mental state, make-believe or imagining, and a special kind of speech act that the author performs in order to transmit that mental state to the reader. Such views have been criticized on the grounds that being performed with intention that the audience should make-believe the content is neither a necessary nor sufficient feature of the author's speech act (Bergman & Franzén, 2022). Relatedly, it has been questioned whether there really is a special kind of mental state, make-believe or imagining, such that it in any way uniquely characterizes our relationship to fiction. If I tell you, in a vivid way, about an adventurous trip I had last year, I might very well do so with the reflexive intention that you imagine how it went down. For make-believe views to successfully shed light on the author's speech act and the notion of truth in fiction, there has to be a state of mind along the of the lines of "imagining-without-believing". Whether there really is such a state remains a contested issue (see e.g. Friend, 2008; Matravers, 2014; Stock, 2017, Chap. 5 for discussion.)

One might be inclined to think that the PP-view, in its appeal to pretence, actually is a kind of make-believe view. Isn't imagining a necessary feature of pretending? Indeed, this is a guiding idea of Walton (1990). If this is right, it might seem that the author's pretended assertions just are invitations to make-believe. But even if a mental state like imagining or make-believe is a necessary feature of pretending (I am agnostic, for the reasons stated above), the explanation of how we read implicit truth into fiction offered by the PP-view is different from that offered by any of the make-believe views currently on the table. On the proposed view, implicit truth in fiction is generated by the author's pretended assertions and the receiver's participation in this pretence. This is what explains the close structural similarity of the implicit story truths to the implied content of the corresponding assertions, as discussed above. None of the proposed make-believe accounts involves any appeals to a pretence by author and receiver, and none of them explains the inferences we make about implicit truth in fiction as structurally analogous to the pragmatic inferences we make with respect to assertions. What we have taken from Walton is the idea of fiction as a mutual game of pretence – a notion that does not play any role for the other Walton-inspired accounts which instead depart from his notion of fictions being "invitations to imagine".

There is a further difference between the PP-view and views that, like Stock's, base truth in fiction exclusively on the intentions of the author. To bring this difference to the forefront, we will consider another class of implicit truths in fiction.

---

[6] This glossing ignores difficulties with unreliable narration. Stock addresses them in a way analogous to the proposal in § 7.

5 Beyond the author

In this section, we will consider cases in which we read more of reality into the fiction than the author or the original readers could possibility have been aware of. Cases like this, I argue, constitute a problem for theories that want to limit truth in fiction to the intentions of the author or the beliefs held by the author or her surrounding society.

Consider first a case from (Franzén, 2021). Tuberculosis is a commonly occurring disease among characters in 19th -century novels. For instance, this is what eventually kills the supporting character Katherina Marmeladova in Dostoevsky's *Crime and Punishment*. Like almost any phenomenon, tuberculosis, as found in the real world, has a number of properties of which a knower at a given time would be unaware. For example, it is today known that the disease is caused by the bacterium *Mycobacterium tuberculosis*, but we can be sure that Dostoevsky did not know this himself, since it was discovered by Robert Koch in 1882, the year after the author's death. In relation to examples such as this, it is, I submit, very plausible to think that when Dostoevsky writes about tuberculosis, he writes about *real* tuberculosis rather than his own representation of tuberculosis. One way of making vivid that the tuberculosis of the fiction indeed must have properties that go beyond any conception of the disease that Dostoevsky or anyone else at the time might have had is to consider whether Katherina Marmeladova, had she lived today, could have been treated with modern medicine. I take it that everyone agrees that she could have. Again, we are not taking Dostoevsky to suggest that Katherina Marmeladova is suffering from a disease that simply shares its name with tuberculosis. On the contrary, we take it that she suffers from *the very same* disease as the one we call "tuberculosis", as made vivid when we ask ourselves how her disease would have been treated today. The fact is that we take it that the tuberculosis from which she suffers is identical to the tuberculosis of our actual world, and this, in turn, means that her illness in the fiction has plenty of properties of which Dostoevsky could not have been aware.

Along similar lines, consider the gold watch that Raskolnikov steals from the pawn-broker in the same novel. Again, it seems very natural to think that the gold out of which the watch is made is actual gold. Dostoevsky is not writing a novel that features some very exotic kind of gold, which is not found in the actual world. To see this, we can ask ourselves whether the watch Raskolnikov stole in the novel would have about the same value as a real 19th -century gold watch. The obvious answer is, I take it, that it would. Again, we have no reason to suppose that the gold of the novel is any different from actual gold. But for the gold of the novel to be like actual gold, and have a similar value, it must again have properties that were undisclosed to Dostoevsky and his original readers.

It is easy to see that the PP-view is in a position to explain these observations in the same way as it accounted for the examples discussed in § 3. The reasoning is similar to that already presented: simply by not mentioning any peculiarities of Katherina Marmeladova's tuberculosis, Dostoevsky has implicated that it is of the normal kind. Since it is, in Levinson's words, "stereotypically exemplified", it carries with it stereotypical properties beyond the knowledge of the author. As already mentioned, generalized conversational implicatures are not elicited through any specifics of the

context, but are rather triggered by default, in the absence of any indication to the contrary.

How does this account compare to how other theories deal with examples of this kind? Let me first briefly mention a strand of theorists not mentioned in the previous section, namely those who aspire to reduce fictional truth to the beliefs of a real or fictional author (Bonomi & Zucchi, 2003; Currie, 1990).[7] To work out what is implicitly true in a fiction, on these accounts, one needs to figure out what the author believes, or, as in Currie's suggestion, the typical beliefs of an average person in the society of origin in the fiction. It should be clear that these theories fail to account for the examples just mentioned.

Next, consider views on which what is true in the fiction is decided by the author's intentions (Byrne, 1993; Stock, 2017). Of these authors, only Stock provides a discussion of how bits of reality are imported into the fiction. According to Stock, while it is not generally the case that fictions include elements from the real world in the way outlined, they might do so in some circumstances:

> […] one way for an author to signal to a reader that she is to imagine that p, given scenario S, is by indicating intentionally that at least with respect to this part of the story, things function as they (commonly are believed to) do in the actual world, where the author also expects the reader to believe that if S was to occur in the actual world, then p would be the case. (Stock, 2017, p. 61)

On the basis of this reasoning, Stock could at least in theory seek to account for our cases by arguing that in these particular examples, the author intends things to work as they do in the real world (including in the bits of reality unknown to the author).

But as Caddick Bourne (2019) points out, Stock does not provide any story about how the author indicates that, for a given part of the story, things function as in the real world. Could Stock avail herself of the same resources as we have and appeal to stereotypicality and similar pragmatic inferences? I think not. The guiding idea behind Stock's framework is "extreme intentionalism", i.e., the idea that all fictional truth is ultimately derived from the intentions of the author. By contrast, it is, as we have seen, a defining feature of generalized conversational implicatures that they are not calculated on the basis of speaker intentions but rather "on general expectations about how language is normally used". As we have seen, this is an important feature of implicatures of this sort: that they can be present even when the implicated content has never crossed the mind of the speaker (in all likelihood, the notion that Holmes might have had three nostrils never crossed the mind of Conan Doyle). It is, I think, an appealing aspect of the PP-view that it combines the broadly speaking Gricean approach to implicit truth in fiction, shared with intentionalist views like Stock's, with allowing for and explaining how truth in fiction can import elements from reality through default inferences in a way that goes beyond anything that has crossed the

---

[7] Lewis' (1978) "Analysis 2" makes truth in fiction dependent on the "collective beliefs" of the society of origin of the fiction and is also subject to these counterexamples. However, Lewis did not endorse it as a universal rule for truth in fiction, but thought that we oscillate between it and the Reality Principle when interpreting fictions.

author's mind. This latter aspect of the theory allows the PP-view to take on board Friend's Reality Principle in a non-ad hoc way.

Stock, of course, would deny that the inference that "things function as they do in the actual world" is a default inference. As indicated in the above quotation, on her view, this inference is valid only in special circumstances. I take it that Stock would simply deny many of the proposed cases of implicit story truth presented above. While I find that denial unintuitive on the face of it, and believe that many people would share my judgment, one should ask if there is some further way of adjudicating the matter.

Stock's proposal seems to predict that there should be at least some possible fictions where what reality is like plays no role whatsoever in the provision of implicit fictional truths. Plausibly, if there are such works of fiction, they should be among those that are set very far from the actual world, such as fantasy stories like *The Lord of the Rings*. In such cases, one might argue, it is so obvious that the author intends the fiction to be radically dissimilar to the real world that we should not expect any overlaps that have not been explicitly communicated by the author. There is no Baker Street in *The Lord of the Rings*, nor is there, as far as I am aware, any documented case of tuberculosis.

But upon further reflection, it is quite easy to see that Tolkien, like any writer, inevitably used building blocks from the real world in creating his fictions. To take a mundane example, it is reported that Frodo is 78 years old at the beginning of the first novel in the *Lord of the Rings* trilogy. So, it is true in the fiction that Frodo is 78 years old. Therefore, it is true in this fiction that the square-root of Frodo's age is about 8.83. Tolkien was, I think it is fair to assume, unaware of this himself. One might think that this is a rather silly point to make. But silly as it may be, it serves to illustrate that even in works of fiction that are far removed from reality, the author inevitably draws from reality and, in doing so, imports facts into the fiction that they are unaware of themselves. A fiction that in no way draws from reality is, on reflection, actually very difficult to conceive of. As the example illustrates, just by making it the case that the characters in a fictional work are of a certain age, many pieces of reality, like the square root of that age, become part of the story truths of that fiction. This makes vivid how difficult it would be to create a story in which *no* elements of the real world are used, and, accordingly, it casts doubt on Stock's contention that background truths of fictions being drawn from the real world are exceptions, derivatives of the author's intention that this should be the case in those particular instances.

This point, that even "outlandish fictions" use building blocks from the real world, is perhaps reminiscent of Hume's contention that, with very few exceptions, our minds seem unable to produce ideas that are not based on prior impressions (Hume, 1999, Sect. 2). With the help of imagination, we can reconfigure ideas in ways that do not correspond to anything real, but ultimately, the material of such reconfigurations will be drawn from outside of our minds. We can make up stories with hobbits and dragons in them, but in doing so, we still use building blocks from the real world.

In summary, then, Stock's view seems to suffer from lack of an explanation of how the author signals to readers that, in these-and-these particular circumstances, they should imagine that things function as they do in the real world. Second, it seems

to wrongly predict that there should be possible fictions where reality is, as it were, completely bracketed.

With all this said, could one not imagine a view otherwise like Stock's, but which drops the commitment to extreme intentionalism and allows for default importation through the mechanism outlined here? Since my proposed view shares with Stock's a broadly speaking Gricean point of departure, the differences between the views would then reside in the intricacies regarding the explanatory appeal to imagining in Stock's view, versus my own articulated in terms of pretence. I will not try to show that the current proposal is superior to such an amended version of Stock's view. What the proposal brings to the table, in comparison with Stock's, is, first and foremost, that it shows how "default importation" from reality is compatible with a Gricean outlook. One can hold that implicit truth in fiction is first and foremost generated through conversational pragmatics, without, as does Stock, implausibly deny what Friend calls the Reality Assumption.

6 The reality principle again

I've compared the PP-view to some competing views and pointed to some advantages that it has over them. But what about the Reality Principle mentioned in § 1, according to which fictional truth is generated on the basis of counterfactuality? Does the PP-view make different predictions about implicit story truth than the Reality Principle, and if so, which view is preferable?

To address this question, let me first mention a particular kind of case that I think is well-accounted for by the Reality Principle but potentially not by the PP-view as currently stated. Such cases concern how the unspoken truths of the fiction *change* (in relation to the real world) based on what is said in the story. For instance, consider a fiction set in late 1815 against the backdrop of Napoleon being victorious at the Battle of Waterloo instead of being defeated. In such a scenario, the reader will naturally infer that Napoleon continues to rule France as emperor and that France remains an empire instead of returning to monarchy if nothing else is indicated. On the face of it, it is not clear that such cases of *counterfactuality* are accounted for by the stereotypicality implicature to which we have previously appealed since involves not just importing elements from the real world, but also making inferences regarding the way unmentioned fact would be *different* from the real world, on the basis of the explicit changes that have been made.

If this is right, how should the PP-view account for examples like this? One way of doing so is by incorporating it as a "third layer" of fictional truth, that is, to hypothesize that, in engaging with fictional discourse, we actually do apply the Reality Principle, but at the stage after pragmatic reasoning has been applied. Our theory of fictional truth would then look like the following:

(1) Fictional truths generated by what is explicitly said by the text.
(2) Fictional truths generated by implied and otherwise indirectly communicated content.
(3) Fictional truths generated by means of letting (1) and (2) form the antecedent of the conditional of the Reality Principle.

As mentioned, Lewis himself might have had this sort of model in mind, as he was happy to include what is "stated *or implied*" (my italics) in "what is told" by the story, i.e., (Lewis, 1978, p.41). While similar to Lewis' (on this interpretation) in accepting this three-layered view of fictional truth, the PP-view differs from his in accounting for a number of cases that Lewis wanted to explain through the Reality Principle by appealing to stereotypicality inferences and similar pragmatic reasoning.

With this refinement of the PP-view at hand, let us now return to the comparison of the current proposal with Lewis' Reality Principle. The current proposal is superior to a theory that seeks to explain all implicit story truths by appealing to the Reality Principle in that also accounts for counterexamples to that theory, like the case with Scrulch the dragon mentioned above. This also goes for a more recent objection to the Reality Principle, raised by Stacie Friend:

> According to Hitchhiker's Guide to the Galaxy, the earth is a supercomputer created by intelligent mice, and an infinite improbability drive allows travel faster than light by passing simultaneously through every point in every universe. Given these primary story-truths, so much would be so radically different from reality that it would seem rash to explain Arthur Dent's reactions to events by appeal to familiar features of human psychology. (Friend, 2017, p. 33)

The idea is that it is true in the *Hitchhiker's Guide* that Arthur Dent's reactions indeed are explained by familiar features of human psychology, but that this is incompatible with the Reality Principle, which, given the strangeness of the universe in which the story takes place, predicts that human psychology should be massively different. However, by taking both what is stated and what is implied as input into the antecedent of the Reality Principle, this objection is avoided. In the novel, it is implied that Arthur Dent's psychology is more or less normal, in the same way as in the examples discussed in § 4.

As examples like this show, the work assigned to the Reality Principle is rather limited in this theory. It is, as it were, applied *after* we enrich the narrator's discourse with pragmatic content and cannot serve to cancel implicatures. Nevertheless, one might still wonder what would motivate a "three-layered" view. It may seem unprincipled to add the Reality Principle as an additional layer, over and above what is stated and pragmatically implicated, even if it's required to account for examples like the one just mentioned. However, the implementation of the Reality Principle in this way interacts smoothly with the basic assumptions we made about fictional discourse in § 2. On this view, again quoting Lewis, "the storytellers pretend to pass on historical information to their audience; the audience pretends to learn from their words, and to respond accordingly." It is natural to take pretending to update one's overall worldview with the new information to be part of pretending to "learn from a narrator and act accordingly". And if one learned about Napoleon being victorious at the Battle of Waterloo, one would conclude that he continued to rule France in the immediate aftermath. That is to say, it's very natural to take pretending-to-learn-from-the-narrator to include making counterfactual inferences about what would otherwise be the case, if what the narrator said was true.

With that said, taking on board the Reality Principle could also open us to the objections raised against it in the literature on truth in fiction (see Woodward, 2011 for an overview). I've brought up two such examples above and explained why the proposed implementation of it is not subject to them. For a more thorough discussion of the objections to the Reality Principle, and an attempt to deflect them, see (Franzén, 2021).

7 From inference to truth

As explained in § 1, the discussion in this text has been conducted under the simplifying assumption that whatever the narrator conveys, explicitly or implicitly, actually is true in the fiction. But as already noted, this is moving too quickly, since the narrator might be misleading us, trying to make us believe something that is not accurate. Again, a typical example is Christie's *The Murder of Roger Ackroyd*, in which it turns out that the narrator himself is the culprit and that he has systematically misled the reader.

This problem is not unique to the PP-view. Anyone who thinks that the narrator's explicit discourse can establish what is fictionally true needs to account for cases when the two come apart. But the phenomenon of misleading narration is, I think, more common when it comes to implicatures. For instance, in Walter Scott's *Ivanhoe*, the narrator talks about a specific character, "The Pilgrim", in a way that falsely suggests that he is not identical to another main character (and that the narrator does not know this), namely Ivanhoe himself. It is never *stated* that the two are not identical, it's just suggested by the way the narrator describes him as mysterious and unknown.[8] More generally, when it comes to the stereotypicality implicatures and similar inferences it might often be that the implicature is simply cancelled, without giving us the feeling of having been misled. For instance, return to Lewis's case with Scrulch the dragon, who we assume can breathe fire even though this is not explicitly stated. It was suggested above that this inference can be accounted for as a stereotypicality inference: it has been implied that Scrulch can breathe fire if it's not explicitly stated that he cannot. But we can imagine that we are not told that many specifics about Scrulch when he is first introduced, but that, somewhere down the line, we learn from the narration that he is not a fire-breathing dragon. In such a case, it would not seem that we have been misled, so much as that we have been provided with more fine-grained information.

In view of these examples, is there nonetheless a way of defining truth in fiction on the basis of what is stated and implied by the narrator? A proposal, that I tentatively endorse is to side-step the issue of unreliable narration by defining truth in fiction on the basis of what has been stated and implied and not retracted or cancelled by the end of the work. Taking into account that we also want to incorporate the Reality Principle in the way outlined above, here is the proposal:

---

[8] Thanks to Andreas Stokke for this example.

(i) **PP-account of truth in fiction**: Where $p_1 \ldots p_n$ is what is pretendedly stated or pretendedly conveyed by the author in a Fiction F, and not retracted or cancelled by the end of the work, it is true in a F that q iff the following holds: were $p_1 \ldots p_n$ the case, q would have been the case.[9]

This formulation is somewhat cumbersome. We should not let that stand in the way of us seeing that the basic elements of this view are rather simple: the author is pretending to tell us something, and we, the addresses, are pretending to take her seriously. What is true in the fiction is what would have been the case if what the author pretends to convey to us, "when all is said and done", really had been the case.[10]

However, D'Alessandro (2016), has raised further skeptical concerns regarding the notion of truth in fiction. He argues that for any work of fiction, we can imagine a sequel to that work, in which some very surprising truths are revealed. A hitherto unknown manuscript could be discovered, in which Holmes does in fact have a third nostril, or putting ourselves in the shoes of the author's contemporaries, a new such manuscript could be produced: "If it could be, or could have been, made true in the Holmes stories that Holmes has a third nostril, then it cannot (now, actually) be true in the stories that Holmes lacks a third nostril, on pain of contradiction." (D'Alessandro, 2016, p. 57).

D'Alessandro's is primarily concerned with arguing for "explicitism", i.e., the notion that only what the narrator explicitly says is true in the fiction. However, as he admits, insofar as the argument is effective, it seems to equally affect explicit story truths (D'Alessandro, 2016, p. 61–63). For any work for fiction, we can imagine a sequel in which something that was earlier explicitly stated is retracted. As an extreme case, think of the infamous "dream season" of the TV-series *Dallas*, at the end of which it is revealed that all events of the season have only been dreamed by one the characters. Explicit story truths are subject to revisability no less than implicit ones.

What I think that examples like this show is not that there is no such thing as truth in fiction, but rather that our hypotheses about what is true in the fiction are in principle always open to revision even to the extent that what was previously true (e.g. that Holmes did not have a third nostril) can be changed.

But suppose D'Alessandro is right, and there is something contradictory in the very notion of revisable truth in fiction. I still don't think that would undermine the current enterprise. Facing the objection that his account is absurd – not going beyond what is explicitly stated is simply not the proper way of engaging with fictions – D'Alessandro argues that explicitism is compatible with that it is still reasonable for the reader to "tentatively suppose" that Holmes's has only two nostrils ((D'Alessandro, 2016, s. 59–60). For the reasons mentioned, I doubt that D'Alessandro's skeptical argument is effective. But even if it is, our theory can naturally be stated in terms of what kind of things readers "tentatively suppose", and the mechanisms behind these supposings,

---

[9] For clarity, this formulation makes it the case that p1…pn will be true in the fiction through triviality. If p1 were the case, p1 would be the case.

[10] "Work" here should be taken to potentially include a series of stories set in what, in contemporary parlor, is called the same "universe".

instead of what is true in the fiction. As mentioned in the introduction, the subject matter of our analysis has been the inferences that we make about what is implicit truth in fiction.

8 Conclusion

I've articulated a theory of truth in fiction (or, if you like, the inferences that we make about it), taking as my point of departure the natural idea that many implicit truths in fictions are conversational implicatures or the result of similar pragmatic reasoning. I've developed this idea through the notion that the author and reader participate in a mutual game of pretence, where the author is pretending to assert and the addressee is pretending to pick up on those assertions. An important feature of the account has been stereotypicality inferences, which I used to explain a range of cases discussed in the literature. More generally, Levinson's neo-Gricean framework allowed for combining a broadly speaking Gricean point of departure, shared with for instance Stock (2017), with an explanation of how important bits of reality are imported into the fiction in a way that can even go beyond the knowledge of both authors and readers, as suggested by Friend's (2017) Reality Assumption concerning truth in fiction.

The general takeaway from the discussion should be that fictional discourse, although pretended, is in fact to a high degree continuous with non-fictional discourse. The inferences we make about implicit truth in fiction are, as I have tried to show, highly similar to inferences about implicatures and other conveyed content that we make regarding non-fictional discourse. To explain implicit truth in fiction, we need not make any assumption about the peculiarities of fictional discourse beyond its role as a part of a game of mutual pretence.

Declarations

References

Bergman, K. G., & Franzén, N. (2022). The force of fictional discourse. *Synthese*, *200*(6), https://doi.org/10.1007/s11229-022-03955-w

Bonomi, A., & Zucchi, S. (2003). A pragmatic framework for truth in fiction. *Dialectica*, *57*(2), 103–120. https://doi.org/10.1111/j.1746-8361.2003.tb00259.x

Bourne, E. C. (2019). Only imagine. *Analysis*, *79*(1), 174–177. https://doi.org/10.1093/analys/any086.

Byrne, A. (1993). Truth in fiction: The story continued. *Australasian Journal of Philosophy*, *71*(1), 24–35. https://doi.org/10.1080/00048409312345022

Currie, G. (1990). *The nature of fiction*. Cambridge University Press. https://doi.org/10.1017/CBO9780511897498

D'Alessandro, W. (2016). Explicitism about truth in fiction. *The British Journal of Aesthetics*, *56*(1), 53–65. https://doi.org/10.1093/aesthj/ayv031

Franzén, N. (2021). Fictional truth: In defense of the reality principle. I E. Maier & A. Stokke (Red.), *The Language of Fiction* (s. 0). Oxford University Press. https://doi.org/10.1093/oso/9780198846376.003.0004

Friend, S. (2008). Imagining fact and fiction. I K. Stock & K. Thomsen-Jones (Red.), *New Waves in Aesthetics* (s. 150–169). Palgrave-Macmillan. https://philarchive.org/rec/FRIIFA

Friend, S. (2017). The real foundation of fictional worlds. *Australasian Journal of Philosophy*, *95*(1), 29–42. https://doi.org/10.1080/00048402.2016.1149736

Grice, P. (1989). *Studies in the way of words*, 1. Print. ed. Harvard Univ. Press.

Hume, D. (1999). *An enquiry concerning human understanding, Oxford philosophical texts*. Oxford University Press.

Kripke, S. A. (2013). *Reference and existence: The John Locke lectures*. Oxford University Press.

Lamarque, P., & Olsen, S. H. (1994). *Truth, fiction, and literature: A philosophical perspective*. Clarendon Press.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

Lewis, D. (1978). Truth in fiction. *American Philosophical Quarterly*, *15*(1), 37–46.

Lewis, D. (1983). Postscript to truth in fiction. I *Philosophical Papers* (s. 276–280). Oxford University Press.

Macdonald, M. (1954). The language of fiction. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *28*, 165–196.

Matravers, D. (2014). *Fiction and narrative*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199647019.001.0001

Predelli, S. (2020). *Fictional discourse: A radical fictionalist semantics*. Oxford University Press.

Recanati, F. (2018). Fictional, metafictional, parafictional. *Proceedings of the Aristotelian Society*, *118*(1), 25–54. https://doi.org/10.1093/arisoc/aoy001

Searle, J. R. (1975). The logical status of fictional discourse. *New Literary History*, *6*(2), 319–332. https://doi.org/10.2307/468422

Stock, K. (2017). *Only imagine: Fiction, interpretation, and imagination* (First edition). Oxford University Press.

Walton, K. L. (1990). *Mimesis as make-believe: On the foundations of the representational arts*. Harvard University Press. https://trove.nla.gov.au/version/20299016

Woodward, R. (2011). Truth in fiction. *Philosophy Compass*, *6*(3), 158–167. https://doi.org/10.1111/j.1747-9991.2010.00367.x