

The fourth wall against possibilism on truth-in-fiction

Deniz Satık, Harvard University
deniz@g.harvard.edu

This paper presents a novel problem for accounts of theories of truth in fiction which hold that fictional entities exist in other possible worlds but not the actual world, following Lewis (1986). This problem is based on the phenomenon in fiction known as *breaking the fourth wall*, in which a fictional character acknowledges their fictionality, either indirectly—by being aware that they are fictional—or directly, by communicating with their actual audience. It is argued that such facts cannot be captured under any version of possibilism. I provide a contextualist account of seeming communication of fictional characters and their actual audience.

Keywords: fiction, possible worlds, fourth wall, contextualism

1 Introduction

It is normal for people to refer to "worlds of fiction" when they discuss the kinds of characters, events, locations and entities that are in a story. For example, in the *A Song of Ice and Fire* series, which the TV show *Game of Thrones* is based on, the world which the story takes place is often called Planetos, which consists of several continents such as Westeros and Essos. The world is greatly enriched with twelve thousands years of history and different kinds of cultures and languages, ranging from the Common Tongue, High Valyrian and Dothraki, that few other worlds of fiction can match. Due to the immense complexity and diversity of Planetos, the author, George R. R. Martin, is often hailed as one of the greatest "worldbuilders" in fiction. Thus, the intuition appears to be that when an author creates a fiction, they are in some sense indicating the fictional worlds which they would like the reader to consider.

Philosophers such as Lewis (1978) have tried to capture this intuition by claiming that true fictional propositions involve facts not in the actual world, but in certain sets of story worlds. For example, possible worlds which correspond to the worlds of *A Song and Ice and Fire* are actually inhabited by individuals like Daenerys Targaryen or her counterparts. Regardless of the metaphysical status of possible worlds, they are by definition are complete and consistent ways the world could have been, and this captures our idea that there are worlds of fiction.

This, I argue, is precisely where possibilist theories of truth in fiction go astray. It is fundamentally difficult for possibilist theories to account for a phenomenon in fiction known as *breaking the fourth wall*, in which characters in fiction acknowledge that they are fictional. As the TVTropes entry on *Breaking the Fourth Wall* meticulously notes, there are many such stories. Beyond this mere acknowledgment, there are even stories in which such characters directly interact with their *actual* audience. This blurs the line between reality and fiction: Ludlow (2006) was the first to note the existence of cross-narrative entities, which can inhabit both real and fictional

worlds. For example, the language *Klingon* was once a fictional entity, but today it has 20 to 30 native speakers according to the Klingon wiki.

My task is simple: I argue that such truths are difficult to account for under theories where worlds are consistent and complete. Under a possibilist theory of truth in fiction, *breaking the fourth wall* phenomena imply characters in fiction think that they are non-actual, which is implausible. I conclude that this problem is unique to possibilist theories, and adopting any other theory of fictional entities would suffice. Furthermore, I provide an account of the seeming communication between fictional characters and their actual audience based on Ludlow (2006)'s application of contextualism to sentences involving truth-in-fiction.

2 Possible world semantics

Possibilism is the view that fictional entities are entities that exist not in the actual world, but in other possible worlds—in other words, they are mere *possible* entities, in the same way that flying pigs exist not in the actual world, but in other possible worlds. On this view, there are possible worlds in which the stories in *Game of Thrones* take place: Tyrion does kill his father with a crossbow while he is sitting on a toilet. Yet, as Kripke (1973) points out, there is more than one possible world in which the stories of *Game of Thrones* take place. There is a world where Tyrion ate oatmeal for breakfast prior to killing his father, and a world where he ate pancakes instead. Which of these worlds do we pick to represent the world of fiction?

The most well-known and developed of this kind is in Lewis (1978), who associates a plurality of possible worlds with the fictional world. Lewis thinks that we should not take the use of fictional sentences such as *Holmes is a detective* at face value; rather, he believes that we should paraphrase them with an intensional operator such as *according to the Holmes stories*.¹ The sentence with the operator is true just in case Holmes is a detective in the worlds picked out by the intensional operator. The sentence *Holmes is a detective* without the operator—in the actual world—is either false or meaningless, depending on one's view of presuppositional failure.

The difficulty, however, lies in picking out the correct set of worlds with the intensional operator. It is not right for the operator to just pick out all the worlds in which all of the events as described by the story are true. It could be the case that our own world is one in which the stories of Harry Potter take place. This is because the wizarding society of Harry Potter keeps itself hidden from "Muggles" such as ourselves through a great deal of concealment magic. Though extremely unlikely, J.K. Rowling could have written true stories of Harry Potter without her knowledge.

Lewis's solution for this is as follows:

(1) **Attempt 1**

"In fiction F, A" is true iff A is true at every world where F is told as known fact rather than fiction. (Lewis, 1978, p. 41)

We should consider the worlds in which events are qualitatively identical to the events as described in the fiction, in which the fiction is told as a story told by a particular storyteller at some time, about known facts rather than about made up stories. But this brings in far too many worlds

¹This is very similar to Walton (1990)'s pretense theory of sentences involving fiction, which proposes that we are engaged in some kind of pretense whenever we utter such sentences. For Lewis, however, the operator is a modal operator that quantifies over fictional worlds.

in which all sorts of things are true. It is intuitively true, but never stated, that Harry Potter is not an alien. Instead, we might pick out the worlds which differ the least from our own world:

(2) **Attempt 2**

"In fiction *f*, *A*" is true iff some *A*-world where *F* is told as known fact differs less from the actual world than any non-*A*-world where *f* is told as known fact. (Lewis, 1978, p. 42)

Once again, this brings in too many worlds. The stories of Harry Potter have ghosts and poltergeists and dragons and fairies, while the actual world likely does not have any of these things. This analysis implies that the people in a story like Harry Potter with these fantastical creatures simply imagine, or even hallucinate, these creatures. In his final attempt, Lewis bases the set of fictional worlds on the prevalent beliefs of the author and his intended audience.²

(3) **Attempt 3**

"In fiction *f*, *A*" is true iff for each collective belief world *w* of the community of origin of *F* some *A*-world where *f* is told as known fact differs less from *w*, than does any non-*A*-world where *f* is told as known fact. (Lewis, 1978, p. 45)

Many shortcomings of Lewis's accounts have been discussed at length in the literature, and several improvements have been proposed. Lewis's account leaves out many worlds in which logically impossible events take place. Currie (1990) observes that a story in which the goal of the protagonist is to refute Gödel's incompleteness theorems makes perfect sense. If the hero was able to accomplish this, this would be true in the fiction, yet mathematical truths are true in all possible worlds. Merely replacing the proof with a surrogate would not suffice, either; the author can insist that the actual proof is the proof the hero seeks to refute.

At the very least, one must expand the possible worlds apparatus by adding in impossible worlds—where logically impossible events happen, or logically impossible objects do exist. This is precisely what Berto and Badura (2019) proposes.

Furthermore, Lewis does not offer an analysis of *metafictional* sentences, which are about truths involving fictional objects in the actual world. Such sentences cannot be paraphrased with an intensional operator:

(4) Harry Potter is a fictional character.

A solution for this problem is provided by Priest (2005) and Berto (2009), who propose that a fictional individual that is a nonexistent entity is present in the actual world, along with existent entities in other possible worlds. For Priest, one could take the nonexistent entity to be the Harry Potter that J.K. Rowling conceives of in the actual world—though he only exists in other possible worlds, not in the actual world. Though Rowling intends a particular nonexistent individual in the actual world, Harry Potter is realized in other possible worlds. This provides a straightforward solution of (4) above.

With the stage set, in the next two sections, I will discuss two problems for the more modern possibilist theories of truth in fiction.

²As Berto and Jago (2013) points out, this analysis has its own problems. What would be the collective belief worlds for a novel written in Nazi Germany by a progressive author who opposed the regime? This would take us out of the scope of the paper, however.

3 The fourth wall

The fourth wall, named after the imaginary wall at the front of a stage which the actors of a play are not supposed to cross, is the wall between the world of fiction and reality. In the vast majority of fiction, the characters are not aware that they are characters in someone else's work of fiction. Yet, there are a great deal of examples in which characters are aware that they are fictional—which is referred to as *breaking the fourth wall*. A famous example of this is Luigi Pirandello's *Six Characters in Search of an Author*. It is a metatheatrical play about the relationship between authors and their characters. The six characters of the play—the Father, the Mother, the Son, the Boy, the Sister, and the Stepdaughter—tell the actors and directors that their author lost interest in them after creating them and that they are in search of a new author.

I am not the first to note cases in which reality and fiction seem to overlap; we discussed the case of Klingon earlier. Woodward (2011) notes John Barth's story *Lost in the Funhouse* in which the character Ambrose comes to believe that he is a fictional character. Though Woodward uses this as an example of an impossible fiction, I do not believe that this is in fact such an example. Ambrose's belief is not a logical impossibility: Descartes himself can attest to the fact that determining whether one exists or not is a non-trivial conclusion. Thus, the following sentence is uncontroversially true, according to Barth's story:

- (5) According to *Lost in the Funhouse*, Ambrose believes that he is a fictional character.

The first, and relatively minor, problem that seems to arise is that of knowledge. The conclusion Ambrose has come to is in fact true. Yet, according to possibilist theories, Ambrose has come to this conclusion in the set of worlds in which he, from his perspective, is actual. In other words, he believes in a proposition that is trivially false, given that he is actual from his perspective. To see why, let us briefly discuss Lewis (1986)'s semantics of actuality.

For Lewis, for a world to be actual does not pick out a special property of that world. It is merely an indexical: in other words, its extension is determined by its context of utterance. The referent of the actual world for any utterance is just the world the speaker is present in, similar to the present tense referring to the moment of utterance. To say that *David Lewis believes that he is a fictional character* would amount to saying that *David Lewis believes that he is non-actual*—a belief that is absurd according to the possibilist. Thus, under the possibilist account, for Ambrose to believe that he is non-actual is implausible: yet, his belief is not only justified according to the story, it is correct. Thus, it appears difficult for a possibilist account to capture Ambrose's belief.³

The basic problem with fourth wall phenomena, however, is more severe than it appears. Let us quickly return to *metafictional* sentences, which as we noted, could be addressed under Lewis's account by assuming Meinongian objects which correspond to fictional characters in the actual world, for metafictional sentences are not preceded by an intensional operator:

- (6) The Mother is a fictional character (in Pirandello's play).

While such a sentence could be true under Priest and Berto's versions of Lewis's theory, it is not possible to account for the truth of such a sentence which is preceded by the intensional operator.

³Adopting an abstractionist approach to possible world semantics, as opposed to Lewis (1986)'s modal realism, would not improve matters here. Even if Ambrose's possible world is not actual—the actual world encompasses all there is—this would not make Ambrose's belief less strange under the possibilist account, given that the only difference is simply that his world does not *obtain*.

As noted prior, Pirandello's story—among many others—smashes through the fourth wall, making the sentence below true:

- (7) According to *Six Characters in Search of an Author*, the Mother is a fictional character.

Though Kroon and Voltolini (2018) notes that usually, sentences like *According to Anna Karenina, Anna Karenina is a fictional character* are false on their de dicto reading, the possibility of such a sentence being true on its de dicto reading is allowed by the breaking of the fourth wall.

The problem, then, for the possibilist is clear. We can paraphrase (7) as follows:

- (8) In the set of worlds corresponding to the story, the Mother is non-actual.

(7) ends up being necessarily false according to the possibilist, even though the whole sentence is true, even according to the story. As noted, actuality for Lewis is merely an indexical—it is necessarily true that the Mother is actual in any world which she is present in. This is no different than a person uttering *I am not at the present time*, which is also trivially false.

One could object that *Six Characters in Search of an Author* is an exceptional story which does not actually involve worlds of fiction, given its focus on metatheatrical criticism rather than worldbuilding. However, there are many other stories in which it is uncontroversial that they take place in a given world of fiction. In Stephen King's *Dark Tower* series of books, which is often hailed for the depth of its worldbuilding, the protagonists meet Stephen King himself, who recognizes them as the characters of in his series of books, while travelling through parallel worlds.

The breaking of the fourth wall, then, is a phenomenon that is ubiquitous throughout all sorts of works of fiction, and it is one that must be taken seriously. The next section discusses a more egregious violation of the fourth wall.

4 Communication between fictional characters and their audience

It is common in fiction for a character to speak to their audience. For example, in the aforementioned *Dark Tower* series, the narrator Stephen King directly addresses the reader just before the ending and tells the reader that the book's ending was fine without an epilogue. More relevant for our purposes is when a character in a given world of fiction directly communicates with his or her audience. Frank and Robin Underwood frequently speak to the audience in the TV series *House of Cards* about their political machinations. Frank Underwood even tells the audience that it is far too late to get squeamish, in regards to the plot.

The following sentence is therefore true with the intensional operator:

- (9) According to *House of Cards*, Robin Underwood spoke to her actual audience.

Under a possibilist theory of truth-in-fiction, this sentence is trivially false, given the analysis of actuality as an indexical. The only possible interpretation of this sentence under which it is true is if Robin Underwood talks to her audience in her *own* world. Yet this is not the interpretation that we have. The watcher's intuition is that Robin Underwood is speaking to the audience directly.

Of course, one could claim that Robin is not really talking to her audience according to the story, and it is merely a narrative device in order for her to clarify her thoughts to the audience.

But there are many stories where this communication cannot be argued to merely be a narrative device. In a theater version of *Peter Pan*, Peter asks the audience if they believe in fairies and to clap their hands to save the fairy Tinkerbell. Peter Pan will save Tinkerbell only if the audience claps loudly enough. Communication with the audience can thus be an essential part of the plot, and hence, the world of fiction.

One way in which the possibilist could try to account for (9) is by assuming that Robin Underwood and Peter Pan do have an audience that they communicate with in their own worlds, and the actual audience merely fills in. But this is completely implausible from the standpoint of fiction, which outright specify the lack of an audience *within* the world of fiction. Robin talks directly to the camera in *House of Cards*: there is no audience to be seen in the Oval Office. Furthermore, given that her machinations are supposed to be a secret, it would not make sense for her to reveal her secrets to an audience in her own world, while keeping it a secret from her enemies!

The way of the possibilist for truth-in-fiction is difficult. The challenge is now to provide an account of (9). What is needed here is a semantics of *speak* that does not imply communication between, for example, actual individuals and Meinongian objects if one adopts a Meinongian view of truth-in-fiction.

One could adopt Walton (1990)'s theory of pretense for truth-in-fiction, which could account for all of the problems noted in this paper. Though I do not wish to dismiss pretense theory outright, Ludlow (2006) notes that Walton (1990) struggles with actually true sentences like (10)-(11) in which fiction and real life seem to overlap—much like the facts proposed in this paper. These sentences involve comparison of fictional characters to actual people:

- (10) Bertrand Russell resembled the Mad Hatter.
- (11) Sherlock Holmes is smarter than any living detective.

It might therefore be wise to consider an alternative to pretense proposed by Ludlow (2006), who notes that sentences like (12)-(13) may be considered true in some contexts but not in others. Ludlow argues that in the context of her playing the role of Buffy Summers, she is not merely playing the role of a vampire slayer or Buffy Summers, she *is* a vampire slayer, and she *is* Buffy Summers. What makes these true, according to Ludlow, are social facts rather than special meta-physical facts:

- (12) Sarah Michelle Gellar is a vampire slayer.
- (13) Sarah Michelle Gellar is Buffy Summers.

Ludlow adopts a contextualist analysis of such predicates, following work in epistemology on contextualism by DeRose (1992), Lewis (1996) and Cohen (1999), among others. The basic idea is that predicates like *know* are context-sensitive in ways that affect their truth-conditions. Whether one's knowledge claim is true could depend on a context of utterance: for example, an academic paper has a much higher standard for truth than a casual conversation at Dunkin' Donuts. A particularly good example of this is provided by Chomsky (1995). Although we would not take a glass of ice tea to be a glass of water at a restaurant, if by some horrible accident our pipes were connected to an ice tea factory and the liquid that came out of our sinks was ice tea, we would still call that liquid water.

This observation applies immediately to the puzzle given in (9) above. *Speak* does not imply direct communication between a fictional entity and an actual one—which would be problematic

for most theories of fiction. In the context of breaking the fourth wall, it merely refers to a fictional character's one-sided attempt to communicate with real-world entities.

5 Conclusion

This paper has argued that the ubiquitous *fourth wall* phenomenon in fiction has consequences on what theory of truth-in-fiction we ought to adopt. My goal here has not been to defend a specific account of truth-in-fiction, however. The problems I have noted here arise solely due to the complete and consistent nature of possible worlds which, according to the possibilist, worlds of fiction are made of. Thus, the solution to this problem is relatively simple, once we give up Lewis's theory. Any other account of truth in fiction, such as stipulating Meinongian objects not just for metafictional truths but for truth-in-fiction cases as well, would be able to address the problems noted in this paper. To conclude, philosophers might benefit from considering the consequences of the peculiarities of fiction, such as the fourth wall, in more depth.

References

- Berto, Francesco. 2009. Modal Meinongianism and fiction: the best of three worlds. *Philosophical Studies* 152:313–334.
- Berto, Francesco, and Christopher Badura. 2019. Truth in fiction, impossible worlds, and belief revision. *Australasian Journal of Philosophy* 97:178–193.
- Berto, Franz, and Mark Jago. 2013. *Impossible worlds*. Oxford: Oxford University Press.
- Chomsky, Noam. 1995. Language and nature. *Mind* 104:1–61.
- Cohen, Ariel. 1999. Generics, frequency adverbs, and probability. *Linguistics and Philosophy* 22:221–253.
- Currie, Gregory. 1990. *The nature of fiction*. Cambridge University Press.
- DeRose, Keith. 1992. Contextualism and knowledge attributions. *Philosophy and Phenomenological Research* 52:913–929.
- Kripke, Saul. 1973. Reference and existence: The John Locke lectures for 1973, oxford university. *Unpublished lectures* Semantics of Natural Language.
- Kroon, Fred, and Alberto Voltolini. 2018. Fictional entities. In *The stanford encyclopedia of philosophy*, ed. Edward N. Zalta. Metaphysics Research Lab, Stanford University, Winter 2018 edition.
- Lewis, David. 1986. *On the plurality of worlds*. Blackwell.
- Lewis, David. 1996. Elusive knowledge 74:549–567.
- Lewis, David K. 1978. Truth in fiction. *American Philosophical Quarterly* 15:37–46.
- Ludlow, Peter. 2006. From Sherlock and Buffy to Klingon and Norrathian Platinum Pieces: Pretense, Contextualism, and the Myth of Fiction. *Philosophical Issues* 16:162–183.
- Priest, Graham. 2005. *Towards non-being: The logic and metaphysics of intentionality*. Oxford University Press.
- Walton, Kendall L. 1990. *Mimesis as make-believe: On the foundations of the representational arts*. Harvard University Press.
- Woodward, Richard. 2011. Truth in fiction. *Philosophy Compass* 6:158–167.