Anti-Realism about Fictional Names at Work: A New Theory for Metafictional Sentences

Louis Rouillé*

Received: 22 February 2020 / Accepted: 2 October 2020

Abstract: In this article, I contribute to ongoing debates about the status of fictional names. The main debate in the philosophy of language focuses on whether fictional names should be thought of as non-referring terms (this is anti-realism) or referring terms (this is realism). This debate corresponds to a debate in metaphysics about the ontological status of fictional characters: the anti-realist claim that fictional characters do not exist while the realist say that they do exist in some sense. Although anti-realism is pre-theoretically intuitive, it has been challenged by a powerful argument in favour of realism based on so-called “metafictional” uses of fictional terms. This argument puts a lot of pressure on the anti-realist, for they have to come up with a theory of metafictional sentences which is in keeping with the anti-realist central tenet. I show that the existing anti-realist account of metafictional statements is wrong-headed. I thus propose a new one. In doing so, I hope to free the anti-realist from the realist pressure. However, I do not offer any argument against realism. Consequently, I merely claim that anti-realism be a live
option. My modest proposal will, perhaps, make anti-realism more attractive than it is today among philosophers of language.

*Keywords:* Anti-realism; fiction; fictional terms; free logic; metafictional statements; reference.

0. Introduction

It is now clear that there are two kinds of philosophers of fiction. Those with the loaded guns are the realist and those who dig are the anti-realist. I am a digger at heart. In this paper, I will present a new direction I consider worth digging to.

Unfortunately, I will not provide a riffle to stir up an anti-realist revolution in the field: I have no argument against realism. I simply hope that by the end of this paper the realist will have new, interesting reasons to put away their guns and come digging with us. It is probable that most will prefer to keep going enjoying their threatening power and rebut at the idea of stooping to the ground. But who knows, they might enjoy the exercise.

1. The present state of the debate between realism and anti-realism about fictional names

I would say that the moment an object appears in a narrative, it is charged with a special force and becomes like the pole of a magnetic field, a knot in the network of invisible relationships. [...] We might even say that in a narrative any object is always magic. (“Quickness”, in Calvino 1988).

Realism about fictional names is the view that fictional names refer, i.e. that fictional names are not empty names. This semantic position entails a metaphysical position called realism about fictional characters, according to which fictional characters exist in some sense. Yet, metaphysical realism about fictional characters should not be thought of as a unified group of

---

1 The entailment is grounded on a principle of compositionality, as made explicit below.
theories. Indeed, metaphysical realists strongly oppose each other when it comes to the fictional entities’ precise ontological status.² The central tenet they share, however, is semantic realism about fictional names.³

Anti-realism, by contrast, is the denial of realism.⁴ Anti-realism thus claims that fictional names never refer, i.e. that fictional names are always empty names. They thus hold that fictional characters do not exist in any sense. There are different versions of anti-realism not so much in the content of the views, but in the way they want to resist realism, as will become clear below.⁵

1.1 The realist guns

Here comes the powerful argument in favour of realism, based on so-called “metafictional uses” of fictional names:

(i) Metafictional statements are truth-evaluable statements containing a fictional name in the subject place.

(ii) The principle of compositionality requires that a name in the subject place of a truth-conditional statement refers.

(iii) Therefore, fictional names refer.

² To name a few: Meinongians like (Meinong 1904) and (Parsons 1980) argue that they are nonexistent objects. (As such, my characterisation of metaphysical realism can be thought of as misleading and this is a reason why I focus on semantic realism below.) Neo-meinongian usually think of fictional characters as abstract objects akin to numbers, for instance (Fine 1982) and (Zalta 1983). Artefactualists, on the other hand, construe them as abstract artefacts, for instance (Kripke 1973/2013), (Van Inwagen 1977), (Salmon 1998), (Thomasson 1999) and (Schiffer 2003). Possibilists view them as concrete unactualised possibilia, for instance (Lewis 1975) and (Lewis 1983).

³ I should say right now that this first presentation of fictional realism is not committing for me, for I will call a “realist” in the rest of this paper anyone who accepts one version of the realist argument given in the next section. As will become clear later on, I focus on different versions of artefactualism stemming from Kripke’s work.

⁴ As such, it can be also be called “irrealism” or “non-realism”.

⁵ Standard arguments leading to anti-realism can be found in (Evans 1982), (Walton 1990), (Everett 2013).
The argument is obviously valid.⁶

Now, here is a metafictional statement which is generally used to load the realist guns:

(1) Emma Woodhouse is a fictional character.

(1) clearly sounds true. Moreover, there is no plausible paraphrase of (1) in which the name “Emma Woodhouse” is not in the subject place.⁷ Therefore, “Emma Woodhouse” refers. So Emma Woodhouse exists in some sense, for it is a fictional character.

1.1.1 On metafictional ammunitions

Metafictional statements are to be sharply contrasted with fictional statements. Fictional statements are statements containing a fictional name and describing the goings on of characters from within their fictional world. In a fictional statement, the name is thus used within pretence or in a game of make-believe⁸ to denote the fictional flesh-and-blood individual. Fictional statements are typically used to express what is true in the fiction. Here is, for instance, the opening line of Jane Austen’s *Emma*:

(2) Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy disposition, seemed to unite some of the best blessings of existence.

---

⁶ Note that the argument is an instance of Quine’s indispensability argument schema, as (Thomasson 2003) rightly remarks. See also (Récanati forthcoming) for a recent, more specific formal rendering of this argument.

⁷ Some people are prepared to come up with such paraphrases using definite descriptions or using quotation marks. This strategy originates in Russell’s work, see for instance (Russell 1919). One can find contemporary views using the same strategy, see for instance (Currie 1990). See also (Dumitru 2015) for an interesting project using a free description theory. I will not follow this line of thinking in this paper though, for I consider that arguments in favour of direct reference for names in general and fictional names in particular are compelling. See in particular (Kripke 1972).

⁸ I take these expressions to be synonymous and use “pretence” systematically to denote the game of make-believe underlying a given fiction.
By contrast, metafictional statements are statements containing a fictional name and purporting to talk about the fictional character from a real-world perspective. The perspective on the fictional character and events is typically external to the pretence. Consequently, metafictional statements are usually not true in the fiction but true simpliciter.\(^9\) Indeed, (1) is not true in Jane Austen’s novel, but it is true at our world. In metafictional contexts, fictional names are not used under pretence but in a serious tone of voice: one can thus define metafictional discourse as “serious discourse with empty fictional names” (Walters 2020, 13). Fictional names, in metafictional contexts, are used to refer to what is ordinarily called “fictional characters” (as opposed to real individuals). To contrast them with the fictional flesh-and-blood individuals, I call them “individuals of paper”.\(^10\)

Given this distinction, one can consider different types of metafictional statements. Here are some which have received a lot of attention in the literature:

(3) Emma Woodhouse is happier than Emma Bovary.

(4) My neighbour is in love with Emma Woodhouse.\(^11\)

Unfortunately, there is no agreed upon exhaustive typology for metafictional statements in the literature.\(^12\) The more systematic attempt, to my knowledge, is to be found in (Woods 2018, 74). It will be useful to rely on his typology for further discussion, for it is quite transparent. I will henceforth say that (1) is a paradigmatic external sentence; (3) is a paradigmatic cross-over sentence; and (4) is a paradigmatic intensional sentence.

\(^9\) There are arguably complications with so-called metafictions, in which the fictional characters are fictional characters \textit{in the story}. See for instance Pirandello \textit{Six characters in search of an author}. I set aside these complications as borderline cases here.

\(^10\) This is a tribute to Plascencia’s metafiction \textit{The People of Paper}, published in 2005. The realists are those who accept individuals of paper into their ontology. However, the ontological status of individuals of paper is highly controversial, depending of each version of realism.

\(^11\) This statement is meant to be akin to statements like “X pities Anna Karenina” or what Oscar Wilde puts into the mouth of Vivian in \textit{The Decay of Lying}:

One of the greatest tragedies of my life is the death of Lucien de Rubempré.

\(^12\) Though some non-exhaustive typologies have been very influential. Especially that of (Van Inwagen 1977) and (Currie 1990).
In principle, any kind of metafictional statement can be used to run the argument in favour of realism.\textsuperscript{13} However, it suffices that one recognises one kind of metafictional statement to run the above argument.

In the following, I will focus on external statements, for they are the least controversial and the more threatening to the anti-realist. However, intensional and cross-over statements will be useful in the following to explain why and how I think the anti-realist have dug their own grave.

\subsection*{1.1.2 The realist program}

The above argument has convinced the majority of philosophers that fictional names sometimes (or perhaps always) refer to individuals of paper. In (Kripke 1973/2013), Kripke forcefully argued that fictional names are “ambiguous”: they do not refer when they occur in a fictional context and they do refer when they occur in a metafictional context.\textsuperscript{14} “Ambiguity”, however, is not the best choice of word. What Kripke meant was that fictional names are polysemous, as is generally agreed upon today, for the two distinct uses of fictional names are systematically related. The polysemy view of fictional names has thus become hugely attractive and many philosophers have subsequently worked on trying to elicit the systematic connection between the fictional and metafictional uses of fictional names. I consider this to be the most influential realist program today.\textsuperscript{15} In the same vein, (Récanati forthcoming) labels the polysemy view the “ecumenical view”, which he defines as “accepting there are two types of use of fictional names, and considering the fictional use as basic”. For this reason, I will

\textsuperscript{13} This point is made in (Everett 2013, 120–38) where he distinguishes between “three forms of argument for fictional realism”. These are versions of the above argument relative to each kind of metafictional statements.

\textsuperscript{14} It should be noted that (Van Inwagen 1977) arrived at a similar view independently.

\textsuperscript{15} Many philosophers thus take the polysemy view for granted. To name a few, see for instance (Dummett 1973), (Searle 1975), (Currie 1990), (Salmon 1998), (Thomasson 1999), (Schiffer 2003), (Braun 2005), (Williamson 2013), (Lycan 2015), (Maier 2017), (Terrone 2017), (García-Carpintero 2019), (Voltolini 2020), (Walters 2020), (Récanati forthcoming).
focus on the polysemy view in the following and see if and how the anti-
realist can resist it.

The core idea of the polysemy view is that the metafictional use “comes
after” or “derives” from the fictional use, in the sense that a fictional
name, by definition, originates in a fiction and can later be used to pro-
duce metafictional statements. From an ontological viewpoint, it is as if
the individual of paper’s coming into existence supervened on the pretence
that there is a flesh-and-blood individual in the fiction. From a semantic
viewpoint, it is claimed that once the pretence is shared and several people
have imagined the flesh-and-blood individual in the way the fiction re-
quired them to, then they can refer back to the individual of paper which
consists in the fictional flesh-and-blood individual “qua fictionally por-
trayed” (Récanati 2018, 10).

There are thus two interconnected theses underlying the polysemy view:
one is a “metaphysical move” during which an individual of paper is “pos-
tulated as a product of [the original] pretence”; the second is a “semantic
move” during which “a fictional name for a person [is transformed] into a
name of a fictional person” (the quotes are in (Salmon 1998, 294) where
Nathan Salmon makes explicit Kripke’s view). The important point is that
polysemy view is essentially dynamic. From a metaphysical viewpoint,
something was brought into existence; from a semantic viewpoint, the fic-
tional empty name metamorphosed into a real name. The acceptance of this
dynamic phenomenon has led to many sophistications which we need not
get into here.16 As will be seen later on, my point of contention with this
kind of realism hinges on a different interpretation of this dynamic phenom-
enon.

16 The most promising sophistication is now Récanati’s theory of dot-concept de-
veloped in (Récanati forthcoming). It combines this research program with the no-
tion of dot-object coming from (Pustejovsky 1998) and reinterprets the notion in the
mental file framework as developed in (Récanati 2012). See also (Terrone 2017) for
an early contribution to Récanati’s view. The reader might like to also consult
(Walters 2020) in which Walters independently develops a view which is inspired by
Pustejovsky’s dot-object theory.
1.2 The anti-realist shovel

Anti-realist, threatened by the realist guns, began to dig. Ironically enough, as I aim to show in this section, the main response was to dig their own grave, without noticing it.

1.2.1 Digging with extended pretences

The main response to the realist argument leading to the polysemy view consisted in denying (i).\(^1\) They said: Although metafictional statements appear to be truth-evaluable, they are not so in fact. Rather, one should construe them as sophisticated fictional statements which are neither true nor false, but fictional.

Of course, metafictional statements cannot be placed down on the same level as their corresponding fictional statements. Indeed (1) is clearly not true in Jane Austin’s novel, nor are (3) and (4). The idea is to defend that they are true in some other relevant fiction. In order to do so, anti-realist have come up with the powerful notion of an “extended” pretence.\(^2\) An extended pretence is a pretence which is parasitic on another pretence, called the “base” pretence. The base pretence corresponds to the original fiction. The extended pretence is a “metafiction”, so to speak: it says how one can talk about the constituents of the fiction, and especially the fictional characters using their names.

To understand how this works, let us look at cross-over statements which are taken to be the most successfully accounted for using this notion of extended pretence.\(^3\) In order to understand (3), one has to merge the two underlying, relevant fictions by Jane Austen and Gustave Flaubert. The merging intuitively consists in having the two Emmas meet in imagination and then compare which is happier, in this imaginative scenario.

---

\(^1\) This strategy was first clearly advocated in the last chapter of (Walton 1990). As mentioned above in footnote, some anti-realists would rather deny (ii) but I think this strategy meets Kripke’s (even more powerful) arguments in favour of direct reference.

\(^2\) The term comes from (Everett 2013) though the notion can be found already in (Evans 1982) and (Walton 1990).

\(^3\) See in particular (Crimmins 1998) for a seminal anti-realist account of cross-over statements.
This extended pretence is a “metafiction” in which both Emmas are characters. By definition of the extended pretence, both Emmas are identical to what they are in their fiction of origin. Given a reasonable principle of reality,\(^{20}\) one understands what “happy” means in this pretence. In particular, since it is generally accepted as true in the real world that one who decides to get married is happier than one who commits suicide, it is fictionally so in the extended pretence. Emma Woodhouse, at the end of Emma, finally decides to get married while Emma Bovary, at the end of Madame Bovary, commits suicide. Hence, it will be fictional in the extended pretence that Emma Woodhouse is happier than Emma Bovary. Therefore, (3) is fictional in the extended pretence. This is why (3) seems to be true simpliciter: because is true in the relevant extended pretence.\(^{21}\)

The anti-realist then need to explain why appearances are deceiving. There are different ways of explaining this.\(^{22}\) A straightforward response consists in remarking that fictional statements are also deceiving. Indeed, metafictional statements should also be deceiving if they are, in fact, sophisticated fictional statements. For instance, (2) seems to be true when compared to:

\[
(5) \quad \text{Emma Woodhouse is an ugly thief who digs her own grave.}
\]

\(^{20}\) It is generally acknowledged that fictions come with so-called “principles of generation” from which one can derive the correct inferences we do in fact derive while enjoying a work of fiction. These warranted inferences define what is “true in the fiction”, or simply fictional. See (Lewis 1978) for a seminal, influential discussion of these. The “reality principle” roughly says that a reader should imagine a fictional world as similar as the real world, unless explicit mention to the contrary. There are many debates about the scope and precise definition of this principle, but virtually everyone agrees that there is always something like a principle of reality when there is a fiction. See (Woodward 2011) for a critical review of the different positions and (Friend 2017) for an influential in-depth analysis of the phenomenon.

\(^{21}\) Note that many realists convinced by Kripke’s argument are ready to accept this. See for instance (Walters 2020, 8) who says that “we can think of [(3)] as being true only within a pretence jointly licensed by the two series of fictions”.

\(^{22}\) These lead to different versions of anti-realism. See for instance the distinction between “radical” and “moderate” anti-realists made in (Récanati forthcoming).
But truth in the fiction is not truth simpliciter. In the same manner, truth in the metafiction is not truth simpliciter. Honor is safe.

At this point, anti-realists have claimed that this strategy can be generalised to all kinds of metafictional statements, provided we can construct a relevant extended pretence. When it comes to external statements like (1) the kind of extended pretence is thus described in (Everett 2013, 65–66):

I suggest that we see the deployment of such expressions as “fictional” as taking place within the scope of an extended pretence. [...] There is the distinction between things that really exist and those which only exist within the scope of the make-believe or fiction. [...] In the simplest cases when we want to articulate [this] distinction, we will engage in an extended pretence in which we pretend that our domain of discourse contains all those entities which occur within some fiction and that those entities are as they are characterised by that fiction. [...] Within this extended pretence, those entities which genuinely exist will count as having the property of being real and those which do not will count as having the property of being fictional.

1.2.2 Why are the anti-realist digging their own grave?

Let us consider a fictional negative existential like:

(6) Emma Woodhouse does not exist.

Fictional negative existentials are negative existentials involving a fictional name. They are good candidates for being external metafictional statements. Indeed, they are clearly not fictional and they closely resemble statements like (1) when it comes to truth-value.

Some may want to doubt that fictional negative existentials are external statements, on the ground that negative existential statements have a controversial logical form anyway. So they would simply dismiss negative existentials from external statements until their logical structure is agreed upon. Indeed, if it turns out that negative existentials are existentially quantified statements involving no name at all, then they fall outside the metafictional data.23

23 In keeping with classic arguments to be found in (Russell 1919) and (Quine 1948).
But this dismissal is, I think, not available to the anti-realist. Indeed, the basic tenet of anti-realism is that fictional characters do not exist (equivalently that fictional names do not refer). So they are, to the bone, committed to the truth of this claim. Therefore, the anti-realist should say that (6) is true. So the anti-realist has compelling reasons to hold that (6) is a good candidate for being an external statement.

On the other hand, the anti-realist cannot deliver truth-conditions for any external statements. This is what has been shown in the previous subsection. The best they can do is to deliver mock-truth-conditions, using extended pretences. Consequently, they are committed to the view that (6) is not, strictly speaking, truth-evaluable.

Fictional negative existentials are thus putting the anti-realist on the verge of inconsistency. The anti-realist is forced to say that (i) fictional characters do not exist and that, strictly speaking, (ii) it is not true that “Emma Woodhouse does not exist”. It should be noted that there is no formal contradiction here because it remains to be shown that general negative existentials formally entail singular ones, which can be resisted.24 This is why I said “on the verge of inconsistency”. However, I think it is fair to say that the anti-realist are forced into a form of theoretical schizophrenia which they should like to avoid if possible.

I should say here that one can find a similar argument against (Everett 2013)’s anti-realism in (Walters 2020, 18). However, it is not quite the same for it relies on a more general argument about Russellian accounts of negative existentials. My argument should thus be thought of as an internal problem for the anti-realist who want to use the notion of extended pretence to deal with fictional negative existentials, hence a somewhat local argument.

1.2.3 Taking stock

Having myself anti-realist intuitions, I can feel an urge to restore a consistent version of anti-realism, regardless of whether this version should be preferable to realism or not at the end of the day. My intuition is that the realist threatening guns made anti-realist dig in the wrong direction. I will therefore advocate digging in another direction.

24 Thanks to Lee Walters for pointing this to me.
The problem here is that the notion of extended pretence is too strong, and should be handled with care. I see no problem analysing away intensional and cross-over statements using extended pretences. However, I think external statements deserve a more careful treatment, as fictional negative existentials show.

My proposal is a kind of divide-and-conquer strategy: first, the anti-realist should divide up the metafictional data into two categories. The intensional and cross-over statements are analysed away using the notion of extended pretence. As for the external statements, the anti-realist should find a way to derive their truth-conditions with the explicit requirement that the fictional name in these contexts does not refer. In order to do this, I use a version of positive free logic. Free logic is not pulled out of a hat and will sound like the obvious response for those who already know about it. Indeed, free logics have been designed to handle both referring and non-referring terms in a single formal apparatus. The difficult part consists in choosing the right free logic and articulating the semantics of external statements with that of the fictional statements and the other metafictional statements in a natural manner: this is the digging part. On this point, my proposal crucially differs from the other available free logic accounts of fictional names, as I will show in due course. Reflecting on my proposal, I will show that it can perhaps be interpreted as a subtle denial of (ii) in the realist argument; I will briefly comment on this in closing this paper.

2. Freedom for anti-realism

2.1 A counter-proposal

The dynamic description underlying the polysemy view of fictional names sounds like magic to me (in Calvino’s sense, from the epigraph of this paper). I think there is an alternative story to tell, which does not rely on any coming into existence of an individual of paper nor on any metamorphoses of names.

What happens is that a new empty name is introduced into serious language. First, a fictional name like “Emma Woodhouse” is introduced within the pretence and everyone using it in such context pretends it refers to the
flesh-and-blood individual in the fiction. Consequently, “Emma Woodhouse” is an empty name because it originates in a fiction and does not purport to name any real individual. Second, one introduces the name “Emma Woodhouse” in the serious language, acknowledging the fact that it is an empty name, alongside with other empty terms like “Vulcan” or “Newman-1”.25 The name “Emma Woodhouse” is crucially different from both “Vulcan” and “Newman-1” because it originates in a fiction, i.e. the first occurrence of the name is within a pretence, whereas the other two have a non-fictional origin. But it is also crucially similar to them in that it is empty. In this counter-proposal, “Emma Woodhouse” is empty all along.

This story is thus in keeping with anti-realism about fictional names. What is true about the polysemy view is that there are two uses of fictional names and that there is dynamic. The metafictional use “comes after” the fictional use. But what is false about the polysemy view is that the dynamic is a metamorphosis of an empty name into a non-empty one. This counter-proposal can be seen as a way of taking the good insights from the polysemy view so as to revitalise anti-realism.

What the counter-proposal needs is a theory which says how one can use an empty name like “Emma Woodhouse” in a subject-predicate statement so as to get an external metafictional statement expressing a true proposition. This is what I will provide now.

2.2 A positive free logic for external metaphysical statements

2.2.1 On the different versions of free logic

Free logic has been designed to handle both referring and non-referring singular terms. The inspiration of free logic was the advent of predicate calculus which was designed to handle both referring and non-referring gen-

25 “Vulcan” was famously introduced by the astronomer Urbain Leverrier to refer to an intramercurial planet which was shown not to exist by Albert Einstein later on. “Newman-1” was introduced by Kaplan in Quantifying in as follows: “I hereby dub the first child to be born in the 22nd century ‘Newman-1’”.

Organon F 28 (1) 2021: 223–252
eral terms, as opposed to Aristotle’s theory of syllogism which cannot handle non-referring general terms. (Lambert 1963) thus presents free logic as an improvement on classical predicate calculus similar to the improvement predicate calculus was on Aristotle’s syllogistic.\footnote{See also (Lejewski 1954) and (Leonard 1956) for previous ideas going in the same direction. However, the term “logic free of existence assumption for singular terms” comes from Karel Lambert, as well as its interpretation as an improvement on classical predicate calculus.}

Such a requirement to handle both referring and non-referring terms\footnote{From now on, I will use the term “term” to denote singular terms only.} on a par, in a compositional, extensional setting (comparable to predicate calculus), however, entails many difficulties. Following a line of thinking originating in Frege’s work, many even argue that this is impossible. They say: Non-referring terms, by definition, do not refer; So they do not have an extension; Therefore, they cannot compose like referring terms. This much is true: the semantic contribution of a non-referring term cannot be of the same nature as that of a referring term. To conclude from this trivial fact that non-referring terms make no semantic contribution is simply incorrect, as free logicians have shown.

Free logicians should now answer the following question: What is the semantic contribution of a non-referring term? The intuitive response is: its lack of referent. Free logic (FL) consists in formalising this idea, so as to integrate it within a compositional, extensional semantic framework delivering truth-conditions.

This idea is challenging because it goes against a core element of extensional semantics, namely that of assigning extensions to both general and singular terms of the language. Indeed, in a classical extensional setting, one uses an interpretation function $I$ to define extensions for predicates and terms such that $I(P^n) \subseteq D$; for every term $t$, $I(t) \in D$. Using this extensional interpretation of the language, one can then recursively define truth-conditions for all formulae in the following manner:\footnote{For simplicity, I will consider a language without identity. Identity is not necessary to handle the linguistic data of this paper.}

\begin{equation}
(\text{atom}) \ v(P^n(t_1, \ldots, t_n)) = 1 \ \text{iff} \ \langle I(t_1), \ldots, I(t_n) \rangle \in I(P^n)
\end{equation}
b) \( v(\neg A) = 1 \) iff \( v(A) \neq 1 \)

c) \( v(A \land B) = 1 \) iff \( \min(v(A), v(B)) = 1 \)

d) \( v(A \lor B) = 1 \) iff \( \max(v(A), v(B)) = 1 \)

e) \( v(A \rightarrow B) = 1 \) iff \( v(A) \neq 1 \) or \( v(B) = 1 \)

f) \( v(\forall x A x) = 1 \) iff for every individual constant \( t \), if \( I(t) \) is defined, then \( v(A[t/x]) = 1 \)

The problematic clause for a FL is (atom). Indeed, in FL, it is not the case that every singular term \( t \) denotes a member of the domain of quantification. In other words, in FL, \( I \) is a partial function which is undefined for the empty terms of the language. (atom) should thus give way to something more complicated in FL:

\[
(FL) \quad v(P^n(t_1, \ldots, t_n)) \begin{cases} \text{(atom) iff } I(t_i) \text{ is defined (for } 1 \leq i \leq n) \\ \text{Something else otherwise} \end{cases}
\]

The next question is: How should we define truth-conditions for atomic formulae when \( I \) is undefined, i.e. when \( t \) is an empty term? (Note that the answer to this question, moreover, needs to be able to feed into the usual inductive truth-conditions for complex formulae expressed in b)-f): this can be seen as a formal constraint.) For instance, take (1) which naturally gets translated in FL as \( F(ew) \). Given that \( ew \) is, \textit{ex hypothesis}, an empty term, how should we define the truth-conditions for (1)? Intuitively, (1) is true. In defining truth-conditions in general, some philosophical choices have to be made and several apparatuses have been designed to implement these choices. The logical space of possible answers defines the different available versions of free semantics.\(^{29}\)

About atomic formulae containing empty terms, there are three positions available. First, \textit{negative} FL has it that such formulae are \textit{always false}:

\(^{29}\) There are many places where one can find a presentation of the different versions of free semantics and a comparison between them. For a seminal account, see (Bencivenga 1986); for a more recent, very systematic presentation see also (Morscher and Simons 2001).
This choice is appealing to some because it takes at face value the metalinguistic biconditional (atom): an atomic formula is true whenever the extensional condition is satisfied (hence when \( I \) is defined) and false otherwise. As such, it is perfectly in keeping with the usual inductive definition of truth-conditions.\(^{30}\) According to (FL–), however, a statement like (1) is false. So it does not fit our purposes.

Second, neutral FL has it that such formulae have no truth-conditions:

\[
\begin{align*}
(FL\#) \quad \forall (P^n(t_1, \ldots, t_n)) & \begin{cases} 
1 \text{ iff } I(t_i) \text{ is defined (for } 1 \leq i \leq n) \text{ and } \\
\langle I(t_1), \ldots, I(t_n) \rangle \in I(P^n) \\
0 \text{ iff } I(t_i) \text{ is defined (for } 1 \leq i \leq n) \text{ and } \\
\langle I(t_1), \ldots, I(t_n) \rangle \notin I(P^n) \\
\# \text{ iff there is a } i \text{ such that } I(t_i) \text{ is undefined}
\end{cases}
\end{align*}
\]

The idea of this position consists in making explicit that empty terms are defective in some sense. The natural way of squaring this position with the usual inductive definition of truth-conditions consists in completing the interpretation function \( I \) in all possible ways and then define a supervaluation function over the set of possible completions of \( I \).\(^{31}\) According to (FL\#),

\(^{30}\) This idea goes back (at least) to the work of the medieval philosopher Buridan (see his *Sophismata – §1.6.5*). The modern motivation comes from the treatment of definite descriptions in (Whitehead and Russell 1912). Negative FL was formalised for the first time in (Schock 1968). Some other influential philosophical motivations for a negative treatment of atomic formulae containing an empty term are given in (Burge 1974).

\(^{31}\) This strategy was first given in (Van Fraassen 1966a) and (Van Fraassen 1966b). See also (Bencivenga 1986) for a now standard version of the supervaluation free semantics which departs substantially from that of Van Fraassen, fixing problems about identity statements involving empty terms.
however, a statement like (1) is truth-valueless. So it does not fit our purposes.

Third, positive FL has it that such formulae are either true or false, for some set of conditions $C$ to be specified:

\[
(FL^+) \ v(P^n(t_1, \ldots, t_n)) = 1 \text{ iff } \begin{cases} I(t_i) \text{ is defined (for } 1 \leq i \leq n) \text{ and } \\
\langle I(t_1), \ldots, I(t_n) \rangle \in I(P^n) \\
C, \text{ otherwise}
\end{cases}
\]

In other words, when an atomic formula contains an empty term, whether it is true or false depends on some explicit condition. It can be seen as designing a semantic module taking care of all and only the problematic formulae, and feeding truth-conditions into the general semantic framework. This position is not a popular choice, for one needs to define explicitly what $C$ is and make sure that these conditions can fit into the usual inductive truth-conditions for complex formulae. My proposal consists in explaining what $C$ stands for here using the counter-proposal I gave above. In order to do this, I will use Antonelli’s formal framework, for I think it is usable as it is.

### 2.2.2 Proto-semantics and fictional terms

In (Antonelli 2000), one can find a bivalent, extensional positive free semantics which is proved to be complete and consistent with the usual inductive definition of truth-conditions. Antonelli’s idea is to introduce a linguistic parameter so as to relativise truth-conditions. The parameter is attached to empty terms, “so that one can speak, in analogy to modal logic, of truth at a term $t$” (Antonelli 2000, 279). Antonelli then formalises this idea using what he calls a proto-interpretation of the language in (Antonelli 2000, 282). The general idea is along the lines given above: one first runs a proto-interpretation which treats empty terms and non-empty terms separately. Then, one feeds the result of this proto-interpretation into the interpretation function so as to finally get truth-conditions.

I will now explain how to define a proto-interpretation using the counter-proposal about fictional terms. Recall, a fictional name like “Emma Woodhouse” first appears in pretence. Within the pretence, the name is a real name referring to the flesh-and-blood individual. It is later introduced in
the serious language as a new empty term. Consequently, the name “Emma Woodhouse” is introduced in the serious language with some information about, at least, its origin. This is the only way we can distinguish between a real name, a non-fictional empty name and a fictional empty name, as discussed above. The set of information accompanying the fictional name should at least contain a mention to the original fiction so as to differentiate a fictional name from a non-fictional empty name like “Vulcan”. This is merely to say that if you did not know the fictional origin of “Emma Woodhouse”, you would mistakenly believe that the name was real and thus treat any sentence containing it as if it were a statement about a real individual. This, I assume, is uncontroversial.

Note that in the picture I am now presenting, one thus needs to have a story about how one extracts the relevant information from a pretence so as to introduce a new fictional name in the serious language. I cannot tell the whole story, though, for it largely exceeds the scope of this paper. However, I think the reader can find such a story in (Evans 1982, 358) when Evans introduces the notion of an “existentially creative pretence”.

Taking Antonelli’s analogy with modal logic at face value, we should think of an empty term as we think of a possible world. We would thus say that a predicate $P$ is true at $t$ iff the predicate $P$ is contained in the informational content of $t$, i.e. the information provided with the introduction of $t$ in the serious language. We would thus write something like: $t \models P$. However, I think this notation is misleading, because one usually writes a proposition in the right-hand-side of a $\models$ and in our context, we talk about predicates. The notation I prefer would rather be something like: $P \in t$. But this notation is also misleading, for it is not true, strictly speaking, that terms are sets containing predicates as elements. Building a set out of a term is the proto-interpretation’s job. A proto-interpretation $\pi$ is a function from terms of the language into sets of predicate such that:

- $\pi(t) = \emptyset$ if $t$ is not an empty term
- $\pi(t) \neq \emptyset$ if $t$ is an empty term

As should be clear by now, when $t$ is an empty term, one should at least find in $\pi(t)$ information about $t$’s origin, in the form of a list of predicates.
We can now explain away condition C and express the truth-conditions for atomic formulae in full generality:

\[(\text{proto-FL}) \quad v(P^n(t_1, \ldots, t_n)) = 1 \iff \begin{cases} 
I(t_i) \text{ is defined (for } 1 \leq i \leq n) \text{ and} \\
\langle I(t_1), \ldots, I(t_n) \rangle \in I(P^n) \\
P \in \pi(t_1), \ldots, P \in \pi(t_n) \text{ otherwise}
\end{cases}\]

One can then define truth-conditions for complex formulae as usual. The application of these truth-conditions is quite straightforward. Let us see how one can derive (1)’s truth-conditions using (proto-FL). \textit{ew} is introduced into the serious language with some information about its fiction of origin, i.e. Jane Austen’s \textit{Emma}. Running a proto-interpretation on \textit{ew}, we should at least have something like: \(\pi(\textit{ew}) = \{\text{empty, origin: } \textit{Emma}, \textit{Emma} \text{ was written by Jane Austen as a fiction, name for a character in that novel, } \ldots\}\) As for the predicate “being a fictional character”, I am not sure how it should be analysed precisely.

Let us suppose, for simplicity, that the expression is something like the conjunction of “being a name for a character” \((N)\) and “originating in a fiction” \((F)\). It seems quite clear that \(N \in \pi(\textit{ew})\) and \(F \in \pi(\textit{ew})\). Therefore, given (proto-FL), it follows that \(v(N(\textit{ew}) \land F(\textit{ew})) = 1\).

Similarly, (proto-FL) predicts that statements like:

\[(7) \quad \text{François Récanati is a fictional philosopher.}\]

are false. Indeed, “François Récanati” is a real name. Hence, in order to evaluate (7), one should look into the extension of “being a fictional philosopher”. Whatever the precise analysis of that natural language predicate, I think one would not find the real individual \textit{François Récanati} as part of the extension of that predicate. Therefore, (7) is predicted to be false. Consequently, the denial of (7) is true:

\[(8) \quad \text{François Récanati is not a fictional philosopher.}\]

I think this squares well with the intuitive truth-conditions of these statements.

To finish with (proto-FL), I think one should remark that \(I(F) = \emptyset\). In other words, in (proto-FL), the extension of the predicate “being fictional”
is the empty set. This means that, properly speaking, nothing is fictional. This is, of course, the central tenet of anti-realism.

2.3 Comparison with other free logic accounts of fictional terms

2.3.1 Negative free logic

The idea of using free logic to model the linguistic data involving fictional names is not new. For instance, one can find an argument in (Evans 1982, 344–9) which establishes very clearly that one is forced to adopt a kind of free logic to deal with (negative) existential statements if one recognises that the word “exist” in these sentences is not merely mentioned but really used. However, Evans did not apply his idea in the subsequent paragraphs, though he indicates that he would go for a negative free logic. I think that something like Evans’s program was taken up and developed by Mark Sainsbury, in a series of publications. Sainsbury’s proposal to use a negative free logic to theorise about fictional discourse is especially developed in (Sainsbury 2007, §6).

My account shares many features with Sainsbury’s account, but it crucially differs in that I use a positive free logic. In particular, I am completely on board with Sainsbury’s departure from the debate about whether one should have a Fregean or a Russellian account of fictional names. What is rightly said about Sainsbury’s proposal in (Orlando 2008, 115) would apply as it is to my proposal:32

[Sainsbury’s] view is, on the one hand, unFregean since it strongly rejects the ascription of descriptive sense to names: names do not have senses which may be construed as ways of thinking about objects, or something along these Fregean lines. On the other hand, it is unRussellian since it does not subscribe to the claim that any genuine name must have a bearer. From Sainsbury’s teleological point of view, the Russellian dictum “Names name” is a generic truth but not a universal one: typically or normally names name but sometimes they might fail to achieve what can be taken to be their proper function.

32 Thanks to an anonymous referee for pointing this to me.
I think, this is typically a view inspired by the reading of Evans which I share. I thus think the debate between neo-Fregeans and neo-Russellians is quite orthogonal to the issue discussed in this paper, namely a theory of metafictional statements, and would use the same arguments which can be found, among other places, in (Sainsbury 2007).

As for the differences, I think there are two points worth commenting upon. First, contrary to Sainsbury, the theory I propose here is a theory of fictional terms and not of empty terms in general. Perhaps my treating fictional names as being introduced in the serious language with some bits of information about their origin can be extended so as to account for non-fictional empty terms like Le Verrier’s “Vulcan”. However, such an extension is not trivial and has not been done here. Consequently, my proposal should be seen as much more modest than Sainsbury’s and somewhat local from a philosophy of language perspective.

Second, I disagree with Sainsbury’s treatment of external statements like (1). According to Sainsbury, (1), when interpreted as a genuine assertion, is false. The reason is simple: according (FL-) atomic formulae containing a empty terms are always false. Sainsbury seems to be attracted to the anti-realist position according to which (1) can be true in some relevant extended pretence, but his position is not so clear from what I could understand. If that is the case, I would say that he is digging his own grave like the other anti-realists.

Here is another way of putting this difference. A good result of Sainsbury’s, however, is that he predicts that statements like (6) are true external metafictional statements. Indeed a statement like:

(9) Emma Woodhouse exists.

is false according to (FL-). Consequently, (6) is true according to (FL-). What I fail to understand in (Sainsbury 2007) is whether there is a logical connection there is between (1) and (6). In my view, it is the truth of (1) which explains the truth of (6). I think this is a good feature of my theory. Such an explanation is not available to Sainsbury, though, for he holds that (1) is false while (6) is true, when both interpreted as genuine assertions.
2.3.2 Positive free logic

Interestingly, there are also positive free logic accounts of statements containing fictional terms in the literature for a long time. The first one I could find is in (Lambert and Van Fraassen 1972, 180–1). Lambert and Van Fraassen’s idea is called “story semantics” and consists in saying that the condition $C$ discussed above is an enrichment of a classical model with a story $S$:

To get all the true sentences in the language we need as part of the model $\mathcal{M}$ also a story. This story has to be consistent with the facts in $\mathcal{M}$, of course; if $\mathcal{M}$ is the real world, the story may say that Pegasus flies, but not that Pegasus exists, nor that Pegasus is identical with some real horse.

Using the contemporary vocabulary introduced, Lambert and Van Fraassen are thus trying to model fictional statements using positive free logic. As I made explicit above, I think this is wrong-headed: fictional statements should not be thought of as genuine assertions but as assertions made within the scope of a pretence. Lambert and Van Fraassen say nothing about metafictional statements.

There is a more recent positive free logic account to be found in (Dumitru 2015) which is, I think, very interesting. The main difference between this proposal and my proposal is that it is taken for granted in (Dumitru 2015, 152) that “fictional terms seem to have a major irreducible descriptive content”. As discussed above with Sainsbury, I am precisely denying this. Dumitru’s account and mine are thus coming from distinct research areas, though they interestingly end up sharing some theoretical commitments.

Moreover, it is possible that another difference between my account and Dumitru’s be roughly the same as the one given above with Lambert and Van Fraassen. Indeed, it seems that Dumitru is not concerned with the distinction between fictional and metafictional statements in his paper. It is true that he calls himself an “anti-realist”, but I am not so sure it

---

33 Thanks to an anonymous referee for telling me this.
corresponds exactly to what I called anti-realism above. Anti-realism, in (Dumitru 2015, 151), is a position in which:

features of fictional objects are ultimately to be explained in terms of features of their marks.

I think this sentence is ambiguous, since “fictional objects” can either mean the flesh-and-blood individuals or the individuals of paper, to use the terminology I introduced earlier. I am not sure how to resolve this ambiguity because Dumitru does not make explicit the precise linguistic data he aims at modelling. If he was to give truth-conditions to fictional sentences using the positive free description theory which he advocates, I would disagree with him on the ground that fictional statements, in my view, do not have truth-conditions for they are not genuine assertions.

3. Anti-realism at works

In this last section, I would like to comment on the status of my proposal to rescue the anti-realist from digging their own grave and finally indicate further research about metafictional statements for the anti-realist.

3.1 Some peaceful reflections about the realist guns

I already said above that the anti-realist strategy I propose is something like a divide-and-conquer strategy against the realist guns. Metafictional statements are indeed problematic if you are an anti-realist. But you should distinguish between the external ones and the others, for the external ones are the more threatening. Then you should deal with intensional and crossover statements using extended pretences and you should deal with the external statements using the positive free semantics I adapted from Antonelli’s proto-semantics.

Now, endorsing free logic, if we reflect on the realist argument once again, can be seen as a way of denying the second premise, namely:

(ii) The principle of compositionality requires that a name in the subject place of a truth-conditional statement refers.
To be more accurate, FL says that this premise is ambiguous, for it really depends on what one means by “a name” here. If it means a *referring term*, then of course the premise is trivially valid, for it goes like this:

The principle of compositionality requires that a referring term in the subject place of a truth-conditional statement refers.

But if it means *either a referring or a non-referring term*, the premise is false, and the argument does not go through. For FL is precisely a compositional apparatus which delivers truth-conditions for statements which contain non-referring terms. I think this points to a subtle interpretation of the notion of compositionality, which is a fundamental notion in the philosophy of language.34

I think it is fair to say that it is a notion of compositionality which does the job in the realist argument. This notion is a complex one, albeit essential to philosophy of language. One part of compositionality, that everyone likes, is the fact that truth-conditions are defined inductively. It is the part which says that the meaning of a sentence is a function of the meaning of its parts as well as the way they are arranged. The second part of compositionality, which is more controversial, concerns how one should give truth-conditions to atomic formulae, those having the form “$S$ is $P$”. Compositionality tells us that such statements are true when the meaning of $S$ and the meaning of $P$ compose in some sense, i.e. relate to each other; the term of art for this basic relation is *predication*.

Frege, who introduced the notion, argued that it is a functional notion: this was a major breakthrough that I think nobody can seriously deny. According to this view, at the core of compositionality, one can find the notion of extensionality. Roughly, what we compose are the extensions which are the domains and co-domains of functions. So the basic blocks of a compositional language should have extensions. Names are one of the basic blocks of language, therefore names have extensions. In the wake of Frege, we thus find this idea that names without extensions are utterly useless, i.e. a dreadful anomaly. Of course, the anomaly is very pervasive in

34 This is not a coincidence, for free logicians are philosophers of language and logicians who have actively taken part in these debates from the fifties on.
natural language, fictional names are but one thorn on the side of Frege’s notion of extensionality.

I think, the subtleness of (proto-FL) as defined above\(^35\) is to show that the problem with empty terms is not with extensionality. Indeed, the proto-semantics defined above is an extensional semantics, handling both referring and non-referring terms. The problem is existence. When a purported referent does not exist, it is not possible to treat it extensionally.\(^36\) The idea is thus to make room for the “anomaly” (like fictional names) while keeping a general extensional framework. Extensionality should thus be restricted hypothetically: the meaning of a sentence can be defined extensionally provided each term has an extension and one should have a back-up plan compatible with extensionality when a term crashes. The anomaly has now joined the rule into big extensional, compositional system. Naturally, the staunch realist will deny this and say that the proto-semantics given above is not extensional, for it has an intentional black box which was obvious in the condition \(C\) given above.\(^37\) I tried to open this black box and show that it is compatible with extensionality. Though I think compatibility is more than enough, some see this as a big let down.

You can now see that I gave what I promised: some new reasons to dig but no rifle to shoot at the realist. I guess I am a pacifist at heart.

### 3.2 Further issues about metafictional statements

Unfortunately, the anti-realist cannot put away the shovel and relax, for there are many open problems in the area which require some more work.\(^38\) I will only mention two big problems ahead. One concerns the delicate problem of so-called co-predication statements like:

---

\(^35\) Probably also of other free logic accounts.

\(^36\) See (Lambert 1981) for a detailed discussion of this.

\(^37\) See (Bencivenga 2006) for an insightful discussion on the distinction between extensional vs intentional systems and their underlying philosophical commitments.

\(^38\) Just like for the realist, by the way, who have a lot of problem solving to do when they stop playing with their guns.
Such sentences share both features of fictional and metafictional statements. Emma Woodhouse here is both a flesh-and-blood individual (she is 21) and an individual of paper (she is the protagonist of a novel). The difficulty is that one should refrain from applying (proto-FL) too quickly. Imagine I accept that some information about the flesh-and-blood individual be stored in $\pi(\text{ew})$ alongside with the information about the origin of the name. Then, where am I to stop? It seems that all the fictional information we have about the fictional Emma can smuggle into the serious language in this manner. Consequently, we will have enough in $\pi(\text{ew})$ to make any fictional statement true simpliciter: this really ruins the whole point of anti-realism.

The problem is thus a delineation problem (which is, by the way, shared by the realist): what counts as fictional information and metafictional information? How should one draw the line? In this paper, I have limited myself to the least controversial bit of metafictional information, i.e. statements like (1). But I have said nothing about how much I am prepared to store in $\pi(\text{ew})$. I am still digging.

The other is the very difficult problem of quantified negative existentials like:

(11) Most of the characters in War and Peace do not exist, though quite a few are historical figures.40

Quantified negative existentials can clearly claim the right to be external metafictional statements, especially if one considers that fictional negative existentials are (as the anti-realist should). After all, the truth of (11) is typically inferred from the truth of relevant (negative) existential statements (“Pierre Bezukhov does not exist”, “Napoléon Bonaparte exists”, etc.). But, the positive free logic I developed is helpless with these statements. Indeed, the semantics of quantifiers was untouched, which means

---

39 This sentence comes from the Wikipedia entry “Emma Woodhouse”.
40 See (Kroon 2003) for an analysis of such statement in an anti-realist frame of mind. Kroon considers this problem the “hardest by far” for anti-realist. See also (Van Inwagen 1977) for an argument against anti-realism using quantified metafictional statements, and (Walton 1990) for a response.
that quantifiers, in free logic in general (thus in (proto-FL)), ranges over all and only the existing individuals (this is clause f) above). So it is impossible to quantify over non-existent individuals in FL. Consequently, quantified negative existentials cannot be given truth-conditions. This is a difficult problem which calls for an extension of positive free semantics so as to account for quantified expressions. Of course, such an extension should not give way to realism, which is a challenge. I am still digging.

Acknowledgments

First I would like to thank Piotr Stalmaszczyk for having organised PHILANG2019 in Łódź where I first presented the ideas one can find in this paper, in a friendly atmosphere. Later on, I developed these ideas in my dissertation which I wrote in the Fall 2019. This paper owes a lot to the heated discussions I had with François Récanati who was my PhD supervisor at the Institut Jean Nicod. I now think that what is presented here is not as incompatible as it first appears with his theory of metafictional statements. Finally, I would also like to thank an anonymous referee for very important comments which made this paper much, much better.

References


Sainsbury, Mark. 2007. Reference without Referents. 2nd ed. Oxford University Press, USA.


