Logic of Existentialism in Fiction

Ilya Makarov,* Valeria Bodishtianu

National Research University Higher School of Economics,
School of Data Analysis and Artificial Intelligence, Moscow, Russia

Abstract

We have considered core approaches to the problem of fictional objects. For each model authors covered the problem whether everything fictional exists or not in terms of evaluation, separating groups of objects, quantifying or existing in modal worlds. The article contains brief overview of the approaches for dealing with fictional objects and evaluating statements containing fictional objects as their part.

1. Introduction

Generally speaking, the problem of non-existent objects has been brought to philosophers' attention several centuries ago. By non-existent objects people understood, firstly, mythical and fictional beings (characters of stories and such), and, secondly, objects, which are rather controversial and impossible in their nature (round square, shapes that do not take space, etc.). In what follows, all the nomenclatures will be developed to define non-existence in all the possible ways. General consensus was that as "existence" can be seen as the property of an object, so "non-existence" could be too.

To understand better what non-existent objects entail, it is important to see the difference between actual and fictional worlds. Fictional worlds are essentially built from scratch by a human mind in the moment of creating a sentence. For example, saying “If I was a girl, I would have studied chemistry” implies the existence of another world S, where aforementioned statement is true. However, the only explicitly told fact about S is that the author is a girl there, which means we have to use out knowledge of the real world to fill in the gaps. The difference from the real-world stories (for example, news articles) is that, firstly, fictional objects and events cannot affect or interact with our world; secondly, we should use implicit knowledge to the smallest degree possible. For instance, from “There is a world with flying pigs” we can assume the appearance of "pigs"; however, we cannot assume that in that world gravitation works the same way, or that people exist at all.

Of the most prominent problems in constructing sentences that make use of nonexistent objects is that we come to the conflict with the idea that for the sentence to be true every atom of it has to be existent.

Therefore, to sum it up, the goal of this research is to find out:
1. Where do the fundamental problems of the logic of fiction lie?
2. Why none of the previous approaches to the problem work in a way we want them to?
How do we handle the property of "existence" in a semantic interpretation of a natural language?

2. The Idea of Nonexistent Objects. Intentionality and Fictional Referrals

One of the founders of the discourse in the logic of fiction and non-existent objects as a whole is Alexius Meinong (Dale 1996). According to him, every thought or even mental activity is directed at some object. One can say that that directness of thought is a referring relation, and, therefore, should abide by the same rules that the referrals of semantics follow.

However, objects that are referred to are not necessary present in our world, material or even possible. We can assign to them whole clusters of properties, most of which are present in real-life objects too. For example, we can consider two objects: Oliver Twist and Donald Trump. Both of them are people, consist of the same materials, and both follow the general rules of human body. Nonetheless, Oliver Twist never really existed in this world materially, so all observations we make about him are deductions in their nature.

Can we really assign the same predicates to both objects, and then somehow assign one to the material world, and another to the fictional one, without anything to distinguish them with? If we do not assign a property or a predicate of existence, sentences “Donald Trump was in London” and

Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
* Corresponding Author
“Oliver Twist was in London” should be treated the same. We'll come to the false conclusion that way, as in case of Oliver Twist we refer to the fictional London; more than that, all the sentences referring to him will generally consist of fictional atoms, which may or may not have their real counterparts in our world. However, if we denote "to be in London" as a predicate $F$, both sentences will have the form of $F(b)$, which clearly is not sufficient.

3. Statements of Nonexistence.
   Russell’s Approach

The problem is on the whole the same: if we say that "Harry Potter doesn't exist" is true, then we assumed that every single part of this sentence is meaningful, or, otherwise, it cannot be true. So, "Harry Potter" is meaningful, and if it is meaningful, it should denote something that exists (Frege 1953).

To overcome this contradiction a descriptivist theory of names was developed in (Russell 1956), which comes down to breaking the object into its descriptors. So, to say that "Harry Potter doesn't exist" we say:

"It's not true that there is a single object, which has the property of being a human, having green eyes, being a character, written by J.K. Rowling, etc."

However, there are several serious problems with that solution. To begin with, many studies argue, that it's basically impossible to include all the descriptors of an object. The cluster of its properties can be complete, but it'll stay infinite. Moreover, the description for a more ambiguous object like God may vary from person to person, and we cannot operate without a universal definition. Secondly, this theory does not work with fictional objects in a way we hoped to; if we say "Aragorn is a good swordsman" we actually mean

$$\exists x : \text{Aragorn}(x) \land \text{good.swordsman}(x)$$

for which the left side is always false, so the whole statement is false. One way to counter it is to make use of "the story-operator strategy".

4. Lewis's Theory

This strategy bids us to rephrase all sentences in regards to the world they exist in. So, we'll have:

"According to Tolkien's "Lord of the Ring" story, Aragorn is a good swordsman."

The bonus of this is that we maintain the rules of classical logic, while still admitting the existence of fictional objects. The main rebuttal was that Lewis considers as truth only what was told explicitly at the story. To circumvent this in (Lewis 1963) Lewis suggested to make use of our own world, so that the statement in the form "In story $S f$" is truth only if it is true in some world, that is the closest to our actual one on the basis of some set of essential characteristics. Basically, we introduce the concept of a "possible" world, which can be obtained from ours through some kind of transformations. After that we have the "population" of our world (the sum of objects), and each object has its own counterpart (or several, according to Lewis) in the other world. The same does not exactly hold true for the object of a fictional world, as there are characteristics that no object of real world holds; however, we can consider their counterpart an object that holds the same properties, but without the impossible one. For example, a talking donkey's counterpart may be simply a donkey, and for any humanoid creature the counterpart will be the human that most resembles it physically.

That approach solves the problem of a dysfunctional body, but I find that the term "closeness" is too loosely determined to use it properly. Nowhere is it stated, what characteristics exactly are considered essential. More importantly, Lewis allows the problem of inconsistency: often in stories the author is too careless, or makes the narrator a careless person intentionally. Lewis decided to allow the "non-consistent truth" in (Lewis 1968). However, later he described a theory, which stated that we can only make logical assumptions using the objects from the same inwardly consistent piece. That makes it rather difficult to operate with other truth from those otherwise consistent fragments, and resembles just forgetting about the problem, rather than solving it.

There is a point that we cannot consider the truth depending on the closeness to our own world; we should consider it depending on the closeness to the $S$ world. Let's examine these two statements:

"In LOTR, orcs have one heart".
"In LOTR, orcs have no heart at all".

If we use Lewis' approach, the statement is true, if it's true in a world closest to our own. As orcs are certainly humanoid creatures, and humans have one heart, we will consider the first statement to be true. However, as it turns out in a different Tolkien's story, the second statement is actually truthful, even though there was no indication of that in LOTR itself. Nonetheless, if we considered the fact that in LOTR world orcs are closer to homunculus in nature than to humans, we would have come to the right conclusion from the start. So the problem of incompleteness remains unsolved.

One other problematic point is that story-operator strategy does not allow the relations between existent and nonexistent objects, and does not help with the objects, belonging to a real world, but not existent, such as some woman's unborn child or a rescheduled meeting.

That problem was first introduced in (Hirst 1991). He says that that for a meeting to be cancelled there has to be an actual object "meeting". However, it does not exist, nor is it fully fictional in our sense. Some can argue that it has
a physical representation – a line in the timetable, someone's notes – however, obviously, if we say:

$$\exists x (\text{represent}(x, \text{meeting}(y)) \land \text{deleted}(x)),$$

we not only once again come to heads with the problem of having no way to interact between existent and nonexistent objects, but we also state only the cancellation of some real counterpart of a lecture (line, notes, etc.), but not the lecture itself.

We can say that the meeting was rescheduled by using the idea that there exists a person that rescheduled it, or:

$$\exists \text{Possible.World} \ (\exists x \exists y \ (x \in \text{Real.World} \land y \in \text{Possible.World} \land \neg (y \in \text{Real.World}) \land \text{person}(x) \land \text{meeting}(y) \land \text{reschedule}(x, y))).$$

But there still remains the problem of quantification over fictional worlds – which states that all fictional objects cannot exist, so our sentence is false by default.

We can consider those objects as objects from the future, which is also not exactly correct (we cannot equalize the idea of lecture $A$ and the factual lecture $A$ that took place in our future world). However, we can try at least to deal with those objects in the same way we deal with the next President of America, or any other future potential object, which simplifies the process (this will we discussed further).

All in all, the idea of possible worlds is still talked about, but without solving the problem of total non-existence and impossibility of quantification over them, it's not sufficient.

Another possible way to overcome the contradiction of nonexistent negations is to simply accept that for an expression to be meaningful it doesn’t have to denote something that exists. That is the route that Meinong chose and developed his whole work on.

5. Meinongian Objects. Parson's theory

Meinongian approach belies on two aforementioned principles: the intentionality thesis and the descriptivist theory (Meinong 1960). Meinong called the collection of all object's properties its "so-being", and every intentional object (that is an object that can be thought about) has a so-being regardless of its existence. Moreover, existence is not a characteristic included in a description, but is a basis for division all the objects into three classes (Dale 1996):

1. Spatio-temporal existence, which is what we all consider a "real" existence.
2. Platonic subsistence, the general idea of something that can still be non-contradictive in real world (fictional characters, ideas like time-machine, etc.).
3. Unreal objects, like a round square, that have the description, but cannot exist without going into conflict with real world's logic.

Generally, the most problematic case (and the one Meinong got the most criticism for) is the last case. It is easy to see that “round square” is a contradictory object. Meinong's answer was introducing the idea that even though we treat nonexistent objects as a different class of existent ones, we cannot apply the same rules of contradiction to them. All in all, he suggested holding quantification only over the possible world domain, which is not very helpful in dealing with fictional objects. Moreover, the impossibility of quantification over nonexistent objects defeats the purpose of the whole logic of fiction, which is to find a way to handle them the same way we do all out other ones.

Meinong continued to develop this theory, but we'll now examine the theory of T. Parsons, who heavily relied on Meinong's theory for basics (Parsons 1980). First of all, Parsons does consider a physical existence a predicate. Secondly, he divides all properties of an object into two groups. First group, nuclear properties, includes all the usual attributes. Being red, round, wet, even being Harry Potter is all a nuclear property. Every object has its own set of nuclear properties, and it's possible to have only one property in a set, even though the resulting object is not necessarily considered existent (for example, "something red").

Second group of properties is extra-nuclear ones. The most oblivious example is, of course, physical existence, but this group also includes the property of being possible, of being created by J.K. Rowling, etc.

One of the problems of Parsons' theory is to determine the line between those two groups. However, later research lead Parsons to the idea of "watering down" some of his extra-nuclear properties to the state of nuclear one, so that people could talk about "existence of round square" without being immediately shot down. In my opinion, that approach only allows to move "impossible" objects to the group of "fictional" ones, which the theory of possible worlds does by itself.

What stayed the same as in Meinong's theory is that each and every world contains every possible object with every set of properties possible. Meinong called that "principle of the unlimited freedom of assumption", and was heavily criticized for using this all-inclusive domain. One of the problems was once again the possible incompleteness of an object: Meinong sustained that it's possible to have an object with only one property, for example, blue. However, according to his own theory, this object also has the property of "having only one property", therefore, it already has two properties, and we come to the contradiction. However, when we consider the fact that "having one property" is now an extra-nuclear one, it is indeed possible to have only blue object. According to Parsons, an object is not created into the universe; it is only a matter of picking it out of the giant pool of all possible objects.

Let's consider the use of Parson's theory in regards to basic axioms of classical logic (Woods 2015):
1. There is nothing whatever that does not exist; or, as it was mentioned by Meinong originally, "There're objects of which it's true that there're no such objects". We will assume that in that language "exist" refers to "being" ($\exists$) and "there is" refers to actual existence ($E_1$), which is a tad counter-intuitive for us, as we are accustomed to the opposite definition, but we'll abide with Woods for now:

$$E_1(x) \land (\exists x).$$

2. No singular referring expression refers unless there exists a referent for it.

$$\forall x (x \rightarrow y) \land \exists y.$$

3. Nothing is something unless it exists.

$$\forall R \forall x (\neg \exists x \rightarrow \neg R(x)).$$

Woods insist on the fourth axiom, the fiction law: the objects of fiction do not exist. He comes from the point that existent and non-existent objects must be separated into different worlds, and objects of fiction cannot in any way be considered "being". However, as we have already seen, that poses a serious problem in describing the relation between those two groups, and also those objects that were or possibly will be in an actual world, and possess all the same properties, but not the "existent" one. Therefore, this axiom for us is a bit redundant already.

The problem with relations between real and non-existent objects is thereby solved for the objects that have the same set of properties. However, there are certain things that are yet to be sorted. What about the fictional objects that do not have the obvious connection to the real world, like a cancelled appointment, or a quenched thought? Those objects are obviously not real, but they still belong to the real world. We are not actually guaranteed to have an object that has the property of being the same, but cancelled, as "cancelled" is an extra-nuclear predicate. For such situations Kripke's system comes to the rescue.

6. KR formalisms

For the most part, using a term in a sentence implies that its denotation does exist, no matter what the term explicitly states. Generally speaking, we need systems that differentiate between "concepts" and "realization of a concept", but treat them with the same capabilities. Problem is, concept is like a subset of a typical Meinongian set of all objects, and so cannot simply be treated as a first-class object. One system that pays more attention to the concepts is SNePS (Shapiro and Rapaport 1986).

SNePS can be considered an "intentional" semantic network, as every bit of information is represented by a node. Arcs in this network represent relations between those concepts. Those nodes can be impossible, nonexistent, not-complete in regards to its set of properties. It allows us to freely represent ideas about concepts themselves and to distinguish assertions and propositions: only a node with no incoming arcs can be considered an assertion.

The problem is, as it follows, SNePS doesn't actually represent real-world extensions or denotations, and it wasn't designed to do so. As SNePS literally shows the world from the mind of an agent, it doesn’t need to distinguish between fiction and reality. Cognitive agent can have two different concepts for the same extension ("Venus" and "Morning Star"), or a concept with no extension at all (impossible objects). To deal with that, creators introduce LEX arcs, where the head of the arc (structured individual node) represents agent’s interpretation (concept) of an actual extension, represented by the tail (sensory node). To sum it up, if $c$ wasn’t previously used, and $w$ is an English word, then a LEX arc from $c$ to $w$ means that $c$ is a Meinongian objects, which corresponds to the utterance of $w$. The same way we can represent the time connection between two nodes, specific properties of an individual object, relation of “being a member of a class”, etc.

However, those arcs seem to be more of a stopgap for the problem of differentiating between intentions and extensions, but still treating them equally. To an outside observer all nodes represent intentions.

In contrast, McCarthy's first-order language unites both intentions and extensions in a single domain (McCarthy 2000). In general there exists a method of connection a concept with its denotations. As we’ve mentioned in the beginning, concept is considered “existent”, if it has at least one denotation. So we make propositions about possible extension of a concept without assuming that all (or any) concepts actually have denotations.

Now, however, we have a polar problem – firstly, concepts do not have any special status to symbolize that they can denote a whole set of objects; secondly, we do not differentiate between kinds of nonexistence in any case.

A certain approach has been suggested in (Hirst 1991), on the basis of Hobbs’ research. It is similar to McCarthy’s language, as no object is assumed to exist unless it is so explicitly stated. Predicate’s arguments can be either existentially transparent (it is implied that the argument is existent) or anti-transparent (object in that position is nonexistent). So, to say that the meeting was rescheduled:

$$meeting(m) \land \exists x(\text{Exist}(E) \land \text{reschedule}(E, x, m)).$$

Predicate “reschedule” would be transparent in its second argument (there indeed exists someone, who rescheduled the meeting), but anti-transparent in the third one (no meeting actually occurred in the real world). Hirst further suggests to differentiate between kinds of nonexistence (add not only a “Fictional” predicate in opposition to “Exist”, but also specify its kind). However, we encounter problems with objects that have interchanging temporal statuses, not to mention the fact that a precise ontology of nonexistent objects is also unclear.

To sum it up, while we do need to treat concepts or intentions as first-class objects, the same as extensions are treated in KR systems, existing approaches do not make it feasible to properly distinguish fictional objects.
7. Modal Logic. Kripke's Worlds

Kripke's semantics is one of the most used systems for formalization Lewis' idea of possible worlds (Kripke 1963). Kripke agreed with the idea of "so-being", so there's a pool of objects that exist in all possible worlds and "real objects" in our notation are just some subset of such pool.

The usefulness of it is that the truth is always relative to the world we operate in. Moreover, in each given world we can operate with not only its domain, but also with every possible object, which solves the problem for crossover relations. Let's consider, for example, the problem of past objects, such as the library of Alexandria. For classical logic and for Russell's approach it's impossible to say "The library of Alexandria was destroyed.", as no such object exists in real life. Now, however, we can say "destroyed (library of Alexandria)", as it's included in the extension of the predicate "destroyed". The same could be said for the problem of crossovers.

On this point though we come to the problem of conscious thought and belief – did I think about the library itself (which existed in the past), that cannot be considered a real object in neither tense? It's agreed that if the thought refers to something that has a real counterpart, we can agree that it refers to the real one, without further confusion. However, it can refer to the real object, but a person may be mistaken about its properties, like "Bill Gates, that guy that developed YouTube". Remembering our problem with crossing in which a fictional object, one that is in our world known as "Aragorn", is not only a good swordsman, but also an existent object, and thus for all the worlds reachable from that one, Aragorn will be treated as such.

That circumvent allows us not to worry about the relations between multiple worlds (as in "book within the book"), but allows to build a depth-tree, the root of which is obviously our own world. To return to the problem of glass and water, we will get:

\[ \Diamond \forall x \ (E! (x) \land water(x) \land myglass(x) \rightarrow wine(x)) \]

It works, because we have a descriptor for our world and the real objects in it, so we do not need to use the substitute from the other world with the same properties.

That leaves us the task of depicting future-changed-in-the-present objects. We can, of course, say, "Hillary Clinton will be the next President of America", because there's a possible world, in which it is year 2020, and there's an existent President of America, and we can work with that. Nonetheless, many people argued that we cannot consider "Hillary Clinton" of our world and "Hillary Clinton" of that fictional world one and the same, even though Kripke's logic explicitly states that this is one and the same Clinton that we take from our pool of objects.

We could see sense in this counterpart theory from the point of completed portrait of Clinton, but if we still allow objects to be seen as the cluster of only the essential properties, it's all good.

8. Cross-world subjunctive modal logic

In (Wehmeier 2012) the author explores the problem of having a predicate occur in a particular world S, and not in an actual world, while still referencing real-world objects and their properties. He makes use of the distinction between "indicative" and "subjunctive" moods for quantifiers. The goal was to express crossover sentences like "If some condition was satisfied, I would have been taller than you". Essentially, he used his mood markers to indicate the world relevant to given predicate, quantifier with its domain, modal operators.

The problem was that the only predicated considered to be "crossover" were binary. There is even an issue of cross-world comparative predicates and, consequently, the author considers degree approach saying "A is higher than B" transforms into "A's height is > B's height", which is not good when it comes to non-quantifiable things like happiness or greatness. We can still say that "I am happier now than I was before", so we have to keep those predicates, but they should refer to objects in question, and not their abstract properties (degrees).

In (Kocurek, 2016) hybrid logic in addition to function approach was suggested. In Kripke models truth of the sentence is always relative to a certain point in a set. In hybrid logic to specify that point we make use of new propositional symbols, nominals, and, consequently, satisfaction operators. Kocurek stated that applying hybrid logic to
function approach could possibly solve the issue with the "actual" function \( \triangleright \), the denotation of an object “as it actually is” in the scope of the currently referenced world.

The problem consists of necessity to either reference several particular worlds simultaneously in regards to different objects, or shift our focus mid-sentence flawlessly. Diagonalization operator \( \downarrow \) helped to a point, as even if we can change the denotation of \( \triangleright \) to the one we need at any point, we cannot shift back to refer to the objects in the real world.

That is where hybrid logic comes in: aforementioned nominals can act as the memory slots that we can “load” at any moment. Shifting between worlds when considering different objects becomes much easier, and we can actually use any kind of a predicate, with no problems that degree approach had. So to say “The yacht could be bigger than it actually is” would transform into:

\[ \downarrow \textit{s.o. Bigger(Yacht, }s Yacht). \]

Even though the problem of crossover relations is rather extensively covered, there remains the question of the impossible objects and our dealing with them, as well as with the worlds that contain them. Moreover, the problem arises when we are trying to reference a fictional object that is not defined in a strictly bounded world, like, for example, Sherlock Holmes. Since in minds of people there exist thousands of versions of a given character, with subtly different characteristics, we cannot express the thought “Sherlock Holmes is more famous than any detective”. Author suggests changing mentions of the “fictional world” to “fictional possibilities”, a set of states.

However, in my opinion, if we consider Parsons’ idea about the universal domain of objects and interchange it with the distinction of domains this approach gives us, we can state that all the versions of Sherlock Holmes that exist denote at the same time that one version of Sherlock Holmes, that contains only explicitly told facts about him in the initial story, as it is in SNePS system, for example. So the previous point would be solved, without once again shifting the mode of operation. The distinction of domains can actually be used to solve the problem Parsons and Meinong had – partial quantification, so as to avoid impossible objects.

Conclusion

We have considered core approaches to the problem of fictional objects. One of them, the theory of "story-operators", severely restricts itself in not allowing the possibility of non-existent objects in our own world.

Everything fictional is said to be a completely separate group. The restriction of predicates in some cases prevents us from describing crossover statements, intentional (known facts about real objects in relation to the fictional ones) statements, inferred (true to the story, but not explicitly told), and many others.

While Lewis did try to include the fictional objects more (which solves a little the problem of inferred and intentional sentences), he still did not solve the problem of the crossover ones. We also have to consider the problem of choosing the essential characteristics for choosing the "closest world" to compensate for the missing characteristics of fictional objects.

Neo-Meinongian approach has more room for development right now. One of its main problems - lack of quantification over fictional domains - is mostly covered by quantified hybrid logic.

Most of the problems that turn up in Russell's approach were covered by having the united domain for all possible objects, and Kocurek gave the necessary tools to somehow restrain that domain. Some other aforementioned conflicts remain unresolved, but it’s still a rather prominent approach.

Acknowledgments

The reported study is to be funded by RFBR according to the research project proposal № 17-01-01534.

References

Parsons, T. 1980. Nonexistent Objects, Yale University Press, New Haven, CT
Woods, J. 2015. Logic and Fiction, in: Handbook of the 5th World Congress and School on Universal Logic, University of Istanbul, 76
Lewis, D. 1968. Counterpart theory and quantified modal logic
McCarthy, J. 2000. First order theories of individual concepts and propositions, Stanford University