Logic and Logical Philosophy
Volume 31 (2022), 579–610
DOI: 10.12775/LLP.2022.020

Newton Peron
Henrique Antunes

A Four-Valued Logical Framework for Reasoning About Fiction

Abstract. In view of the limitations of classical, free, and modal logics to deal with fictional names, we develop in this paper a four-valued logical framework that we see as a promising strategy for modeling contexts of reasoning in which those names occur. Specifically, we propose to evaluate statements in terms of factual and fictional truth values in such a way that, say, declaring 'Socrates is a man' to be true does not come down to the same thing as declaring 'Sherlock Holmes is a man' to be so. As a result, our framework is capable of representing reasoning involving fictional characters that avoids evaluating statements according to the same semantic standards. The framework encompasses two logics that differ according to alternative ways one may interpret the relationships among the factual and fictional truth values.

Keywords: philosophy of fiction; fictional names; logic of fiction; four-valued logics

1. Introduction

Proper names that purport to refer to fictional characters (henceforth, *fictional names*) pose some difficulties when it comes to presenting a logical theory that countenances them. In this paper we discuss what some of those difficulties are and develop a four-valued logical framework that we see as a promising strategy for representing contexts of reasoning in which fictional names occur. Specifically, we propose to evaluate statements in terms of factual and fictional truth values in such a way that, say, declaring 'Socrates is a man' to be true does not come down

Special Issue: Logics and their interpretations. II. Guest Editors: H. Antunes and D. Szmuc

Published online May 25, 2022

to the same thing as declaring 'Sherlock Holmes is a man' to be so — for while the former is true according to the way the real world is, the latter is true only in virtue of what take place in certain piece of fiction. As a result, our framework is capable of representing contexts of reasoning that involve statements in which both fictional and non-fictional names occur while avoiding that those statements be evaluated according to the same semantic standards.

The paper is structured as follows: in Section 2 we outline some linguistic intuitions that motivate our proposal and which highlight some limitations we have identified in the classical, free, and modal logic approaches to modeling reasoning in contexts in which fictional names occur. Those intuitions guide our subsequent choices concerning the formal language and the two formal semantics to be presented in Sections 3 and 4. In Section 5 we present one system of natural deduction rules for each of the semantics in Section 4, and prove a few results about them. In Section 6 we then prove that both systems are sound and complete. Finally, in Section 7, we indicate how the two systems presented in the paper could be modified to deal with inconsistent and incomplete fictional scenarios, along with other possibles lines of further research.

Before getting started, though, it is worth making two caveats about the methodology and scope of our proposal. The first one is that we take here a moderate pluralist stance on logic. We do not claim that our approach is the only way of formally modeling the rules people follow when reasoning about fiction. We are prepared to admit that in different contexts we can and do use quite different rules (and therefore different logics) to reason about fiction. That being said, we do not wish to deny that the classical, the free, or the modal logic approches provide us with useful models for representing reasoning in which fictional names occur. Nonetheless, each of those approaches does have its limits, and what we want to do in this paper is to *explore* what we take to be a promising and interesting alternative that is capable of overcoming or avoiding at least some of them.[1]

The second caveat is that this is a work on logic rather than on metaphysics. Our interest is to offer *one* way of formally representing the rules people seem to follow when they reason about fiction in at least

[1] Indeed, we acknowledge that our approach also has its own stock of limitations, which are pointed out in Section 7 and in a few footnotes. We expect to be able to deal with them in future works, by further developing the framework presented here.

some circumstances. Thus, in spite of the importance of those questions, we are not concerned with whether fictional objects exist, and if they do, what their nature is. Nonetheless, we are also well aware there is no way of conducting investigations on the logic and the semantics of fictional discourse that is completely neutral with respect to metaphysical issues. After all, we are here aiming towards what Kripke [14] calls an *applied semantics*, as opposed to a pure semantics. And, for this reason, our proposal *is not* free of metaphysical commitments. This is a fair concern, and one we fully acknowledge. Hence, although we will inevitably make some metaphysical assumptions in this paper, we will not take a definite stance about them, since they are only the by-products of our semantical and logical choices.

2. Objections to the classical, free, and modal logic approaches

If we take fictional names to be formally represented by singular terms, it's natural to consider them to be terms with no reference (i.e., as *empty terms*). In classical logic, this approach leads immediately to some very undesirable results, though. For consider the sentences:

$$\text{Sherlock Holmes is a detective} \tag{1}$$

$$\text{Sherlock Holmes is a superhero} \tag{2}$$

If we let the individual constant s be a formal name of Sherlock Holmes, and the predicate letters D and H be the formal rendering of respectively 'is a detective' and 'is a superhero', then (1) and (2) will be formalized as Ds and Hs. From either sentence we are authorized to infer $\exists x(x = s)$, which in classical logic is the standard formalization of:

$$\text{Sherlock Holmes exists} \tag{3}$$

(3) seems false, though, at least in some contexts in which the word 'exists' is used. Furthermore, many people that accept that Sherlock Holmes does not exist would be inclined to admit that (1) is a true sentence — in some sense of 'truth' — while (2) is false. Hence, in classical logic from a sentence that many consider to be true we may infer a sentence that many consider to be false.

This result is a consequence of the fact that in classical logic every term refers to some individual in the domain of quantification and that the existential quantifier $\exists$ is understood as carrying ontological commitment by the vast majority of classical logicians. Now, logicians have come up with different strategies that were intended to accommodate non-referring names within the classical framework. One such strategy is to represent empty names as singular terms but assign them a bogus referent in the domain of quantification. In the case of fictional names, for instance, a certain first-order language may include singular terms for both 'Sherlock Holmes' and 'Odysseus', but require them to refer to a certain arbitrary object, say, the empty set. Accordingly, every singular term will then refer to some object or other in the domain of the quantifiers.[2] It is quite clear, however, that the theory does not provide a satisfactory solution to the problem above, for not only would the formal rendering of (1) and (2) fail to express (1)'s and (2)'s intended meaning, but also make both false, given that the empty set is certainly not a detective. Moreover, any existential sentence, such as (3), will come out true, contrary to our intuitions.

Taking fictional names to be represented by singular terms is not the only option available to the classical logician, though. According to Russell, 'Sherlock Holmes', along with the vast majority of proper names that occur in natural language, is not a genuine name, but rather certain definite description in disguise. And given Russell's general view that definite descriptions are expressions to be paraphrased away in every sentential context in which they occur, (1), (2) and (3) should be analyzed as:

$$\text{There is exactly one } x \text{ such that } x \text{ is} \ldots \text{and } x \text{ is a detective} \tag{1r}$$

$$\text{There is exactly one } x \text{ such that } x \text{ is} \ldots \text{ and } x \text{ is a superhero} \tag{2r}$$

$$\text{There is exactly one } x \text{ such that } x \text{ is} \ldots \tag{3r}$$

Here, '. . . ' abbreviates all (or at least the most salient) characteristics of Sherlock Holmes as described in Conan Doyle's novels. Since there exists no (single) individual which satisfies all those characteristics, (1r), (2r) and (3r) are false, and so are (1), (2) and (3). In addition to yielding intuitively wrong results (e.g. in the case of (1)), this descriptivist account also faces similar difficulties whenever *mixed statements* are concerned — i.e., statements that are about both real and fictional individuals. For

[2] Logicians have made use of this strategy for different purposes, such as, for example, defining the abstraction operator $\{x : \alpha(x)\}$ as a total functor in set theories based on ZF [see 25, §2.5].

example, the sentence:

$$\text{Conan Doyle created Sherlock Holmes} \tag{4}$$

is to be paraphrased as:

There is exactly one x such that ... and x was created by Conan Doyle (4r)

And since (4r) is false, so is (4), which, again, also defies our semantic intuitions.

Sainsbury [22, ch. 6] argues for the adoption of a *negative free logic*. Positive and negative free logics are commonly based on a bivalent semantics. That is, analogously to classical logic, any formula is assigned exactly one of two truth values. However, they differ from classical logic in the way they handle sentences in which empty terms occur (i.e., terms that either have no referent at all or refer to an object outside the domain of quantification). In negative free logics, all atomic sentences with empty terms are invariably false, while in positive free logics, some of those sentences may be true.[3]

According to Sainsbury, negative free logic offers a much more natural account of sentences in which fictional names occur than classical logic does, for in a negative free logic (2) is false in all models in which s is an empty term. The downside of this approach, however, is that (1) is also false in those models, offending the semantic intuition of many people, who tend to admit (2) as being true compared to (1). It seems, as has been suggested by Orlando [20], that any view that entails that (1) is false is as unsatisfactory as any view that entails that (2) is true.

Perhaps, the adoption of a *positive free logic* would be a more promising choice, for unlike negative free logics, they do not rule out true atomic sentences in which empty terms occur. Accordingly, there are models in which (the formal counterpart of) (1) is true, even though both (2) and (3) are false. These models are such that: the interpretation of s belongs to the extension of D; it does not belong to the extension of H; neither does it belong to the domain over which the quantifiers range. As a result, we can construct a model in which Ds is true, while Hs and $\exists x(x = s)$ are both false, as expected.

Despite some advantages of adopting a positive free logic over a negative one [see, e.g., 2], this strategy suffers from what we see as some serious expressive limitations. First, in a positive free logic one is unable

[3] See [4, 16, 18] for comprehensive overviews of free logics.

to express general statements about fictional objects, such as:

$$\text{Sherlock Holmes arrested some criminals} \tag{5}$$

In the context of classical logic, where the quantifiers range over absolutely all individuals, (5) could be expressed by:

$$\exists x(Asx \wedge Cx) \tag{5f1}$$

(where A is the formal counterpart of the predicate 'x arrests y', and C of the predicate 'x is a criminal'). However, since in positive free logics the quantifiers are only supposed to range over non-fictional objects, (5f1) does not retain (5)'s intended interpretation, for what (5f1) actually expresses in such a logic is that Sherlock Holmes arrested at least one real criminal, which is clearly untrue.[4]

The second problem of adopting positive free logics for modeling reasoning about fiction is that although they allow for true atomic sentences about fictional objects, they do not recognize any differences between truth assignments to sentences in which fictional names occur and to sentences in which they do not occur. Consider, for example, the sentence:

$$\text{Dave Toschi is a detective} \tag{6}$$

and compare it with (1) above. It is clear that whenever one declares (1) and (6) to be true, the relevant senses of 'true' in each case are significantly different. For while (1) is true only according to what is told in the stories of Conan Doyle, (6) is true in virtue of *the fact* that Dave Toschi, the main investigator in the Zodiac Killer case, is a real-world detective. In a positive free logic, however, the truth of (1) and (6), when translated into the formal language, would come down to exactly the same thing, viz., the satisfaction of certain predicates by certain individuals. As far as we can tell, neither the semantics nor the syntax of a positive free logic have the necessary resources for appropriately expressing this difference.

One way out would be to take into account the semantic notion of a possible world. Thus, in the actual world (6) is true while (1) is false.

[4] One might retort that the introduction of substitutional quantifiers Π and Σ, in addition to the usual objectual ones (viz., $\forall$ and $\exists$), could overcome this difficulty. For, unlike $\forall$ and $\exists$, Π and Σ could be taken as ranging over absolutely all individuals. However, this move would only work if every individual had a corresponding name in the formal language, which is at odds with the fact that several novels contain characters that do not have names.

But there is a possible world where (1) is true which, of course, is not the actual one.

The semantics of a quantified modal logic may be formulated using constant or variables domains. Now, the proposal of applying constant domain semantics to fictional names is subject to the following criticism: if Sherlock Holmes exists in a possible world, he exists in the actual world; and if Dave Toschi exists in the actual world, he exists in any other possible world. Hence, all individuals that exist in some world necessarily exist, which is not intuitive at all.

Much more promising is the modal logic with variable domains. Thus, in Conan Doyle's fictional stories, the individual Sherlock Holmes exists, but not in our actual world. Davi Toschi, in turn, exists in our actual world but does not exist in the world created by Conan Doyle. Every fiction here is seen as a possible world, and every possible world has its domain of objects existing in that possible world.

Modal logic with variable domains delivers a useful and interesting framework for dealing with fictional entities, but it also has its limits. Note that in modal logic every individual is a possibly existing individual. This doesn't seem to be a problem in the case of Sherlock Holmes. From (1), we would infer something much weaker than (3). We would simply infer that it is possible that Sherlock Holmes exists. The main problem is that in some fictional texts contradictions appear, so that there are individuals who are logically impossible, for example in [7, 21].

It is true that this criticism can be avoided through a semantics of impossible worlds, as advocated in [3], for instance. The impossible worlds would be those in which contradictions occur and, therefore, the individuals that exist in that world would not be possible in our world, which we could assume as a consistent world.

Anyway, we would like to add a second criticism here to modal logic with variable domains that somehow also affects approaches with impossible worlds too. Note the following sentence:

$$\text{Superman is a superhero} \tag{7}$$

Basically, the same people who are willing to admit (1) as being true would admit (7) as being true as well. On the other hand, compare the following sentences:

$$\text{It's possible that Sherlock Holmes exists} \tag{8}$$

$$\text{It's possible that Superman exists} \tag{9}$$

Most people would be inclined to reckon (8) as true while very few would have the same attitude towards (9). This is because there are several different notions of ‘possible’ in natural language. Due to Superman’s supernatural properties, he doesn’t seem to be a possible individual, at least physically or biologically so. Sherlock Holmes, on the other hand, has no supernatural properties, making it much easier for us to admit that he is a possible individual.

One could, of course, reply that the semantics of possible worlds is rich enough to express different nuances of the notion of ‘possible’ according to the relations of accessibility between possible worlds: logical possibility, physical possibility, metaphysical possibility, etc. This can be seen in the huge number of distinct modal systems existing in the literature, many of them seeking to capture different nuances of the meaning of ‘necessary’ and ‘possible’ in natural language. Our argument does not depend, however, on the distinct meanings of ‘possible’ in natural language. Rather, it runs as follows: if some people can admit (1), (7) and (8) as being true while (9) as being false, this seems to indicate that the possibility or not of the existence of fictional terms in some usages seems irrelevant to admit the truth or falsity of sentences in which fictional terms occur. And this is reason enough to admit that whereas modal logic with variable domains offers a rich and expressive formal semantics for dealing with fictional terms, it also has its limits.

3. Two four-valued approaches

The foregoing discussion brings us to the following intuition: although it seems that some sentences about fictional beings and events, such as (1), are true, the sense of ‘is true’ at stake in these cases is significantly different from the sense in which sentences about real beings and events are so.

The most intuitive notion of truth is, of course, somehow dependent on how the world really is, and if non-fictional names are the only names that occur in a sentence, then the sentence as a whole should be considered as either true or false according to the facts that obtain in the world. On the other hand, a quite different notion of truth seems to be presupposed whenever fictional statements are concerned, since their truth depends for the main part on what is told by or depicted in one or another fictional media. Accordingly, we propose that the notion of truth (falsity) be split into two: *factually true* (*false*) and *fictionally true* (*false*). Given an atomic sentence Pt, where P is a predicate and t a term, we may then say that Pt is factually true (false) if the referent of t does factually exist and is such that it belongs (does not belong) to the extension P; and that Pt is fictionally true (fictionally false) if t does not refer to anything that factually exists and is such that t belongs (does not belong) to the extension of P. By evaluating sentences in this way, we are capable of counting (1) and (6) among the truths, while still acknowledging the difference between the senses of 'is true' at stake in each case: while (1) is fictionally (though not factually) true, (6) is factually (though not fictionally) true. Likewise, the sentence:

$$\text{Sherlock Holmes is not a detective} \tag{10}$$

which is the negation of (1), is fictionally (though not factually) false, but the negation of (6):

$$\text{Dave Toschi is not a detective} \tag{11}$$

is factually (though not fictionally) false.

Notice that to say that (1) is not factually true does not entail that it is factually false (in addition to being fictionally true) and that to say that (10) is not factually false, does not entail that it is factually true (in addition to being fictionally false). Hence, the values *factually true* and *factually false* are *not* mutually exhaustive, as neither are *fictionally true* and *fictionally false*. This means that some sentences may be understood as lacking a truth value whenever we restrict our attention to their factual or fictional values. For example, (1) lacks a factual truth value, while (6) lacks a fictional truth value.[5]

Evaluating sentences according to whether they have a factual or a fictional truth value introduces some new challenges when it comes to determining the truth conditions of complex sentences. To understand

[5] Nonetheless, the systems to be presented in the next sections allow neither for truth values gaps nor for truth value gluts (in the sense that sentences must receive at least one of the four values, and that each value excludes all the others). We are aware that this aspect represents a limitation of our proposal. After all, not only does the sentence 'Sherlock Holmes owned six pipes throughout his life' lack a factual truth value, but it also lacks a fictional value. In future works, we intend to overcome this limitation by modifying the systems to make room for gaps and gluts (see Section 7 for further comments on this point).

how these challenges emerge, consider the sentence:

$$\text{Sherlock Holmes is a philosopher and Socrates is a detective} \tag{12}$$

(12) is false since both conjuncts are false. But is it factually or only fictionally false? As we shall now see, either answer can be considered correct depending on how one chooses to interpret the behavior of the logical connectives and quantifiers concerning the interaction of factual and fictional truth values. Since we have as yet found no definite reason for adopting one interpretation over the other, we shall not take a stance about this question in this paper, but merely indicate how those two options can be implemented in the corresponding formal semantics.

According to the first interpretation, (12) is factually false, given that at least one of its conjuncts, viz., 'Socrates is a detective', is factually false. The reason is that, on this interpretation, truth values are ordered as follows:

$$\textit{factually false} < \textit{fictionally false} < \textit{fictionally true} < \textit{factually true}$$

And since we take conjunctions to have the least value among the values of their conjuncts, this means that because 'Socrates is a detective' is factually false, so is (12). Likewise, the disjunction:

$$\text{Sherlock Holmes is a detective or Socrates is a philosopher} \tag{13}$$

is factually true because at least one of its disjuncts, viz., 'Socrates is a philosopher', is also factually true (no matter the value of the other disjunct). According to the second interpretation, on the contrary, (12) is rather only fictionally false, given that at least one of its conjuncts, viz., 'Sherlock Holmes is a philosopher', is fictionally false. Thus, on this interpretation, truth values are ordered as follows:

$$\textit{fictionally false} < \textit{factually false} < \textit{factually true} < \textit{fictionally true}$$

And, as the reader might expect, (13) will then be fictionally true, given that 'Sherlock Holmes is a detective' is so.

As a matter of fact, not only are there two possible interpretations for evaluating complex formulas, but a similar divide also shows up even at the atomic level. Take, for instance, Sainsbury's example in [22, ch. 6]:

$$\text{Tony Blair admires Coriolanus} \tag{14}$$

and suppose that (14) is true. Since Coriolanus is one of Shakespeare's characters, should (14) be factually or fictionally true? One might maintain that (14) is factually true because it talks about a real person and expresses a fact in the real world. But (14) might also be taken to be only fictionally true, since 'Coriolanus' is a fictional name, and so there is nothing in the real world itself that could be the fact of Coriolanus being admired by Tony Blair.[6] As before, we shall not take a stance about which of these two ways of interpreting (14) is indeed correct, but, as in the case of complex formulas, merely describe how these alternative interpretations may be implemented in an appropriate formal semantics.[7]

Allowing sentences to be either factually true (false) or fictionally true (false) seems to call for some way of expressing this difference within the object language itself. In other words, we need some formal analogue of the natural language phrases 'it is factually true that', 'it factually holds that' etc. To accomplish this, we shall make use of a new primitive unary sentential operator ! such that whenever ! is prefixed to a formula α, it expresses that 'it is a matter of fact, rather than of fiction, whether α', while $\neg!\alpha$ expresses that 'it is a matter of fiction, rather than of fact, whether α'.[8] As result, we will be able to express the following four

---

[6] A third position is to maintain that it is not possible to decide whether (14) is factually or fictionally true. However, even in a non-deterministic approach we must follow, for technical reasons, the principle expressed in [1] that non-determinism takes place only concerning non-atomic formulas. We will return to this point in the Section 7.

[7] Lehman's [15] *strict Fregean logic* is intended to be a logical rendering of Frege's views on the semantics of fictional names in [8] as those views are accounted for by the orthodox interpretation of Frege's work on the philosophy of language. According to this interpretation, though fictional names, such as 'Odysseus', have a sense, they do not have a corresponding referent; and to the extent that sentences refer to a truth value (either *the True* or *the False*) only if all terms occurring in it are non-empty, any sentence in which fictional names occur will lack truth value.

Nonetheless, some scholars have also advocated other, less orthodox, interpretations of Frege's views about the semantics of fictional names. By drawing on [9, p. 130], for example, Evans [6, pp. 291–321] maintains that according to Frege fictional names have both fictional senses *and* fictional references. If Evans' interpretation is correct, then sentences in which at least one such name occurs should refer to a fictional truth value. This is precisely what takes place in the second interpretation of (13) above, in which it is enough for a fictional name to occur in an atomic sentence for it to have a fictional truth value. Thus, if Lehman is indeed entitled to call his logic a *strict* Fregean logic, we also have good reasons to believe that the logic **L1** below is a *broad Fregean logic* — since it is a formal counterpart of a less unorthodox interpretation of Frege's views about fictional names.

[8] We chose of the symbol ! as a loose reference to the expression **E**!, which is sometimes used as the existence predicate of some free logics. Though the logics to be

scenarios

- α is factually true: $\alpha \wedge !\alpha$
- α is fictionally true: $\alpha \wedge \neg!\alpha$
- α is factually false: $\neg\alpha \wedge !\alpha$
- α is fictionally false: $\neg\alpha \wedge \neg!\alpha$

Now, consider the following sentence:

It is factually the case that Sherlock Holmes is a detective (15)

which can be formalized as:

$$Ds \wedge !Ds \tag{15f}$$

Everyone will agree that (15) (and so (15f)) is false. But should it be factually or only fictionally so? We consider it quite reasonable to admit that whenever assertions to the effect that something is *factually* true or false are concerned, we are interested in what takes place in the real world, rather than what is told in some fictional story. Hence, ! will be interpreted here in such a way that $!\alpha$ can never receive a fictional value: if α is either factually true or false, then $!\alpha$ is factually true; and if α is fictionally true or false, then $!\alpha$ is factually false. Notice that according to this interpretation, any iteration of ! $(!!\alpha, !!!\alpha, \dots)$ will always be factually true, no matter the truth value of α.[9]

Let us consider now how to express existential claims such as (3). We have already seen that classical logicians face the dilemma of either taking sentences such as (1) as true, which appears to entail (3) to be false, or else embrace the consequence that both (1) and (2) are equally true. But either option leads to very counter-intuitive results. Free logicians are no better off, for not only do they face some of the problems classical logicians do (such as taking (1) to be false if the free logic in question is negative) but also seem utterly unable to express certain general statements about fictional characters (e.g., (5)). Our proposal is vulnerable

presented here also have a primitive existence predicate in their logical vocabularies, we shall adopt the symbol **E** instead of **E**! (which is also common in free logics).

[9] ! bears some important resemblances to da Costa's [5] consistency operator $^\circ$, which identifies, within the object language, those formulas that have "a classical behavior" (viz., that obey the principle of explosion: $\alpha, \neg\alpha \vdash \beta$). Similarly to $^\circ$, ! indicates in the object language whether a certain formula α has a factual truth value. Thus, we may say that as α° expresses that α has classical behavior, $!\alpha$ expresses α has a "factual behavior".

to none of the above problems, though. For consider (3) again. Is it factually false or fictionally true? Well, the answer depends on whether we take (3) to express a fact about the real world, or merely describe what takes place in Conan Doyle's novels. For if we are entitled to claim that (1) is fictionally true because Sherlock Holmes is a detective according to Doyle's stories, there is no reason why (3) cannot be also fictionally true, given that he also exists according to those very same stories. In order words, (3) may be formalized in either of the following two ways:

$$\exists x(x = s) \wedge !(\exists x(x = s)) \tag{3f1}$$

$$\exists x(x = s) \wedge \neg!(\exists x(x = s)) \tag{3f2}$$

depending on whether (3) is taken to express that Sherlock Holmes factually exists or exists only according to a certain piece of fiction. But, on this interpretation, the inference from (1) to (3), when rendered in either of the logics to be presented in Section 4, turns out not to be valid. For even though $\exists x(x = s)$ does hold in both logics, in order to obtain (3f1) or (3f2) we still need an additional premise, viz., either $!(\exists x(x = s))$ or $\neg!(\exists x(x = s))$, neither of which is itself valid.[10]

That fictionally true existential claims may be expressed in terms of the existential quantifier means, *contra* Quine, that $\exists$ is not an ontologically committing expression — in the sense that if $\exists x\alpha$ is true, then there *factually* exists an individual that satisfies α. Hence, even though $\exists x\alpha$ does entail real existence if $\exists x\alpha$ is factually true, $\exists$ can no longer be taken as a *univocal* way of expressing ontological commitments. To fill in this gap, and also for technical reasons which will become clear below, we shall further enrich the formal language with a new unary predicate **E**, the *existence predicate*. For every term t, **E** is meant to indicate whether t denotes a real object, as opposed to a mere fictional one. Thus, unlike $\exists x(x = t)$, $\mathbf{E}t$ expresses that t factually exists, and so *is* ontologically committing. The relations among !, $\exists$, and **E** can be

[10] If s denotes a fictional individual, (3f1) is factually false in either of the formal semantics presented in Section 4. However, (3f2) will be assigned different values. In **L1** (3f2) is fictionally true, given that $\exists x(x = s)$ is so (no matter the fact the $\neg!(\exists x(x = s))$ is factually true), while in **L2** it will be factually true, given that $!(\exists x(x = s))$ is also factually true. Since (3f2) is intuitively fictionally true, this result might favor the adoption of **L1** over **L2**. However, we do not think this to be a knock-down reason for rejecting **L2** altogether, since we acknowledge that there might be other ways of expressing existential claims in which the second (non-ontologically committing) sense of (3) turns out to be fictionally rather than factually true in **L2**.

summarized in the following two equivalences which hold in both logics to be presented below:

$$\mathbf{E}t \text{ is factually true iff } !(\exists x(x = t)) \text{ is factually true}$$
$$\neg\mathbf{E}t \text{ is factually false iff } \neg!(\exists x(x = t)) \text{ is factually false}$$

Finally, it is worth mentioning that by having both ! and **E** at our disposal, we are capable of formally expressing the two possible interpretations of (13) mentioned above. In the first interpretation, which maintains that (13) is factually true because it talks about at least one real object (viz., Tony Blair), every n-ary predicate P is such that $!Pt_1 \dots t_n$ and $\mathbf{E}t_1 \vee \dots \vee \mathbf{E}t_n$ are equivalent; while in the second interpretation, according to which (13) is merely fictionally true because 'Coriolanus' is a fictional name, the corresponding equivalence holds between $!Pt_1 \dots t_n$ and $\mathbf{E}t_1 \wedge \dots \wedge \mathbf{E}t_n$. As we shall see briefly, while the first equivalence is a primitive rule of the logic **L2** to be presented in Section 5, the second equivalence is a primitive rule of the logic **L1** instead.[11]

In the next sections, we will develop two four-valued logical frameworks that are based on the conceptual discussion carried out in this section.

4. Syntax and formal semantics

The logical vocabulary of the two logics to be presented in this and in the next section is composed of the sentential connectives $\neg$, $\wedge$, and !, a denumerable set $\mathcal{V} = \{v_n : n \in \mathbb{N}\}$ of individual variables, the universal quantifier $\forall$, the predicates $=$ and **E**, and parentheses. We shall specify the non-logical vocabulary of any first-order language by means of its *first-order signature*, i.e., a triple $\langle \mathcal{C}, \mathcal{F}, \mathcal{P} \rangle$ such that the elements of $\mathcal{C}$, $\mathcal{F}$, and $\mathcal{P}$ are respectively the individual constants, function symbols, and pred-

[11] The reader might wonder why we don't allow predicates to be divided between factual and fictional, just as terms are. The reason is that we can easily express the distinction as follows:

$$Px \text{ is a factual predicate iff } \forall x!Px \text{ is factually true,}$$
$$Px \text{ is a fictional predicate iff } \forall x!Px \text{ is not factually true.}$$

This may be an interesting to the realism/antirealism dilemma of fictional properties posed by Sawyer in [24].

Obviously, our language cannot determine whether fictional properties exist, for to do so we would need a second-order language and, in that case, we would have to abandon our completeness result.

icates letters of $\mathcal{S}$.[12] The *existence* and *identity predicates*, $\mathbf{E}$ and $=$, are assumed to be among the predicate letters of any first-order signature.

Given a first-order signature $\mathcal{S}$, the sets $Term(\mathcal{S})$ and $Form(\mathcal{S})$ of respectively the *terms* and *formulas* of $\mathcal{S}$ are inductively defined in the usual manner, except that the definition of $Form(\mathcal{S})$ also has the following additional clause: if $\alpha \in Form(\mathcal{S})$, then $!\alpha \in Form(\mathcal{S})$.[13] We take the connectives $\vee$, $\rightarrow$, and $\leftrightarrow$, and the existential quantifiers, $\exists$, to be defined as usual: $A \vee B := \neg(\neg A \wedge \neg B)$, $\alpha \rightarrow \beta := \neg\alpha \vee \beta$, $\alpha \leftrightarrow \beta := (\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$, $\exists x A := \neg\forall x \neg A$.

As for the formal semantics, an interpretation of a first-order language is a structure in the usual model-theoretic sense, but with two (disjoint) domains rather than a single domain. The two domains are supposed to represent respectively the set of all real and the set of all fictional individuals.[14]

Definition 4.1. Given a first-order signature $\mathcal{S}$, a *structure* for $\mathcal{S}$ is a triple $\mathfrak{A} = \langle A, A_\emptyset, ^{\mathfrak{a}} \rangle$ such that: (i) A and $A_\emptyset$ are respectively the set of real and the set of fictional individuals of $\mathfrak{A}$, $A \cap A_\emptyset = \emptyset$, and $|\mathfrak{A}| = A \cup A_\emptyset \neq \emptyset$ ($|\mathfrak{A}|$ is the *domain* of $\mathfrak{A}$); (ii) $\mathfrak{a}$ is a function (the *interpretation function*) satisfying the following conditions:

- For each individual constant c of $\mathcal{S}$, $\mathfrak{a}(c) = a$, for some $a \in |\mathfrak{A}|$;
- For each n-ary function f of $\mathcal{S}$, $\mathfrak{a}(f)$ is a function $g\colon |\mathfrak{A}|^n \rightarrow |\mathfrak{A}|$;
- For each n-ary predicate P of $\mathcal{S}$, $\mathfrak{a}(P) \subseteq |\mathfrak{A}|^n$.
- $\mathfrak{a}(\mathbf{E}) = A$ and $\mathfrak{a}(=) = \{\langle a, a\rangle : a \in |\mathfrak{A}|\}$

From now on, we shall write $f^{\mathfrak{A}}$ and $P^{\mathfrak{A}}$ instead of respectively $\mathfrak{a}(f)$ and $\mathfrak{a}(P)$. If $\mathfrak{A}$ is an $\mathcal{S}$-structure, then an *assignment* in $\mathfrak{A}$ is a function

---

[12] Each function symbol and predicate letter of $\mathcal{S}$ is assumed to have a corresponding finite arity.

[13] Henceforth, we shall use x, y, z as variables ranging over $\mathcal{V}$, lower-case Greek letters, α, β, $\gamma, \ldots$, as variables ranging over $Form(\mathcal{S})$, and upper-case letters, Γ, Δ, $\Sigma, \ldots$, as variables ranging over sets of formulas. In order to improve readability, we adopt the usual conventions concerning the omission of parentheses. The notions of a *free* and a *bound occurrence* of a variable in a formula, as well as the notions of a *term free for a variable* in a formula, *closed term*, and *closed formula* (or *sentence*) are as in classical logic. The set of closed formulas and closed terms over $\mathcal{S}$ will be denoted by $Sent(\mathcal{S})$ and $CTerm(\mathcal{S})$, respectively. Finally, we will write $\alpha[x/t]$ to denote the formula obtained by replacing every free occurrence of x in α by the term t.

[14] This assumption is far from consensual. Many may assume, for example, that the real New York City is the same as the fictitious New York City in the Spide-rman stories. We will continue with this controversy in Section 7.

$s : \mathcal{V} \longrightarrow |\mathfrak{A}|$. For each assignment $s \in |\mathfrak{A}|^{\mathcal{V}}$, and term t, the *denotation* of t in s (denoted by $\overline{s}(t)$) is defined as usual: (i) if t is a variable x, then $\overline{s}(t) = s(x)$; (ii) if t is an individual constant c, then $\overline{s}(t) = c^{\mathfrak{A}}$; and (iii) if t has the form $ft_1 \ldots t_n$, then $\overline{s}(t) = f^{\mathfrak{A}}(\overline{s}(t_1), \ldots, \overline{s}(t_n))$. If $a \in |\mathfrak{A}|$, we shall denote by s_x^a the assignment that is just like s except that it assigns the individual a to the variable x.[15]

Given an $\mathcal{S}$-structure $\mathfrak{A}$, each formula will have one (and only one) of the following four values in a structure $\mathfrak{A}$ (with respect to an assignment s): T (for factually true), t (for fictionally true), f (for fictionally false), and F (for factually false). There will be, however, two alternative ways of assigning those values. Each of these ways corresponds to one of the alternative interpretations of mixed sentences discussed in the last section. This divide will be present in all semantic clauses, and so, for each logical symbol, we shall present the relevant semantic clauses at once, identifying them with the numerals '1' and '2'.

Let P be a predicate letter distinct from $\mathbf{E}$ and let i range over $\{1, \ldots, n\}$. Then:

(**P1**)

- $(Pt_1 \ldots t_n)^{\mathfrak{A}}[s] = T$ iff $\langle \overline{s}(t_1), \ldots, \overline{s}(t_n) \rangle \in P^{\mathfrak{A}}$ and $\overline{s}(t_i) \in A$, for every i;
- $(Pt_1 \ldots t_n)^{\mathfrak{A}}[s] = t$ iff $\langle \overline{s}(t_1), \ldots, \overline{s}(t_n) \rangle \in P^{\mathfrak{A}}$ and $\overline{s}(t_i) \in A_{\emptyset}$, for some i;
- $(Pt_1 \ldots t_n)^{\mathfrak{A}}[s] = f$ iff $\langle \overline{s}(t_1), \ldots, \overline{s}(t_n) \rangle \notin P^{\mathfrak{A}}$ and $\overline{s}(t_i) \in A_{\emptyset}$, for some i;
- $(Pt_1 \ldots t_n)^{\mathfrak{A}}[s] = F$ iff $\langle s(t_1), \ldots, \overline{s}(t_n) \rangle \notin P^{\mathfrak{A}}$ and $\overline{s}(t_i) \in A$, for every i.

(**P2**)

- $(Pt_1 \ldots t_n)^{\mathfrak{A}}[s] = T$ iff $\langle \overline{s}(t_1), \ldots, \overline{s}(t_n) \rangle \in P^{\mathfrak{A}}$ and $\overline{s}(t_i) \in A$, for some i;
- $(Pt_1 \ldots t_n)^{\mathfrak{A}}[s] = t$ iff $\langle \overline{s}(t_1), \ldots, \overline{s}(t_n) \rangle \in P^{\mathfrak{A}}$ and $\overline{s}(t_i) \in A_{\emptyset}$, for every i;
- $(Pt_1 \ldots t_n)^{\mathfrak{A}}[s] = f$ iff $\langle \overline{s}(t_1), \ldots, s(t_n) \rangle \notin P^{\mathfrak{A}}$ and $\overline{s}(t_i) \in A_{\emptyset}$, for every i;
- $(Pt_1 \ldots t_n)^{\mathfrak{A}}[s] = F$ iff $\langle \overline{s}(t_1), \ldots, \overline{s}(t_n) \rangle \notin P^{\mathfrak{A}}$ and $\overline{s}(t_i) \in A$, for some i.

[15] Of course, whenever $s(x) = a$, then s and s_x^a will not differ at all.

According to (**P1**), an atomic formula α is either fictionally true or fictionally false if and only if some of the terms occurring in α are fictional — or, equivalently, α is factually true or factually false if and only if all the terms occurring in α are non-fictional. On this interpretation, mixed formulas are therefore always fictionally evaluated. Conversely, (**P2**) entails that α is either factually true or factually false if and only if some of the terms that occur in α are non-fictional, which corresponds to the idea that mixed atomic formulas should always be evaluated in terms of the factual truth values, T and F.

Unlike the other predicates, the existence predicate **E** will be assigned a special role in the formal semantics, for, as we have seen, it is meant to allow us to distinguish terms according to whether they denote real or fictional entities. Since claims to the effect that a certain term is either fictional or non-fictional are factual, $\mathbf{E}t$ should never receive either the value t or f. As a result, we end up with the following semantic clauses for atomic formulas formed by **E**:

- $(\mathbf{E}t)^{\mathfrak{A}}[s] = T$ iff $t \in A$; and
- $(\mathbf{E}t)^{\mathfrak{A}}[s] = F$ iff $t \in A_{\emptyset}$.

As with the predicate symbols, we also have two ways of interpreting the sentential connectives. The relevant semantic clauses may be summarized in the following tables:

α	$\neg\alpha$
T	F
t	f
f	t
F	T

	$!\alpha$
T	T
t	F
f	F
F	T

$\wedge_1$	T	t	f	F
T	T	t	f	F
t	t	t	f	F
f	f	f	f	F
F	F	F	F	F

$\wedge_2$	T	t	f	F
T	T	T	f	F
t	T	t	f	F
f	f	f	f	f
F	F	F	f	F

Remark 4.1. Notice that, as in Gödel's $\mathbf{G}_4$ and Łukasiewicz's $\mathbf{Ł}_4$, conjunction behaves as the minimum operator in the truth value orders of both **L1** and **L2**. Negation behaves as in Łukasiewicz's systems. Specifically, if x and y are truth values, then $x \wedge y = \min\{x, y\}$ and $\neg x = 1 - x$.

The disjunction and implication operators of **L1** and **L2** are classically defined — i.e., $x \vee y = \max\{x, y\}$, as in **G4**, and $x \to y = \max\{1 - x, y\}$. Thus, the implication operators of **L1** and **L2** differ from those of **G**4 and $\mathbf{Ł}_4$.

It is worth emphasizing, however, that **L1** and **L2** have two designated truth values, which is in stark contrast with the systems in Gödel's and Łukasiewicz's hierarchies [see 10].

Finally, taking a to range over $|\mathfrak{A}| = A \cup A_{\emptyset}$, we have the following clauses for $\forall$, which correspond to the two interpretations of $\wedge$ presented above:

(**Q1**)

- $(\forall x\alpha)^{\mathfrak{A}}[s] = T$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] = T$, for every a;
- $(\forall x\alpha)^{\mathfrak{A}}[s] = t$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] \in \{T, t\}$, for every a, and $(\alpha)^{\mathfrak{A}}[s^a_x] = t$, for some a;
- $(\forall x\alpha)^{\mathfrak{A}}[s] = f$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] \neq F$, for every a, and $(\alpha)^{\mathfrak{A}}[s^a_x] = f$, for some a;
- $(\forall x\alpha)^{\mathfrak{A}}[s] = F$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] = F$, for some a;

(**Q2**)

- $(\forall x\alpha)^{\mathfrak{A}}[s] = t$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] = t$, for every a;
- $(\forall x\alpha)^{\mathfrak{A}}[s] = T$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] \in \{T, t\}$, for every a, and $(\alpha)^{\mathfrak{A}}[s^a_x] = T$, for some a;
- $(\forall x\alpha)^{\mathfrak{A}}[s] = F$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] \neq f$, for every a, and $(\alpha)^{\mathfrak{A}}[s^a_x] = F$, for some a;
- $(\forall x\alpha)^{\mathfrak{A}}[s] = f$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] = f$, for some a;

From the semantic clauses above and the definition of $\exists$ in terms of $\neg$ and $\forall$, it is straightforward to obtain corresponding clauses for the existential quantifier. In the case of **L1**, for example, we have:

- $(\exists x\alpha)^{\mathfrak{A}}[s] = T$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] = T$, for some a;
- $(\exists x\alpha)^{\mathfrak{A}}[s] = t$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] \neq T$, for every a, and $(\alpha)^{\mathfrak{A}}[s^a_x] = t$, for some a;
- $(\exists x\alpha)^{\mathfrak{A}}[s] = f$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] \in \{F, f\}$, for every a, and $(\alpha)^{\mathfrak{A}}[s^a_x] = f$, for some a;
- $(\exists x\alpha)^{\mathfrak{A}}[s] = F$ iff $(\alpha)^{\mathfrak{A}}[s^a_x] = F$, for every a.

Notice that the characterization of a first-order structure in Definition 4.1 bears some resemblance to the corresponding definition in some positive free logics, in which interpretations also have two domains, D_i and D_o, called respectively *inner* and *outer* domains. D_i is usually taken to represent there the set of individuals that exist, while D_o is a superset of D_i that also includes the "individuals" to which empty terms (e.g., 'Pegasus') refer. Thus, the domains A and $A_{\emptyset}$ in Definition 4.1 correspond respectively to D_i and D_o/D_i. There is, however, a crucial difference between our approach and the one based on dual domain free logics, for while the quantifiers $\forall$ and $\exists$ range over only the individuals in D_i, the quantifiers above range over *the entire domain* $A \cup A_\emptyset$ of $\mathfrak{A}$. This decision corresponds to the view, discussed in Section 3, that $\forall$ (and thus $\exists$) is an ontologically neutral expression, something which most (if not all) free logicians would certainly reject.[16]

We shall call the logics generated by the two interpretations above **L1** and **L2**, respectively. Given $\Gamma \cup \{\alpha\} \subseteq Form(\mathcal{S})$, α is a *semantic consequence* of Γ in **L1** (**L2**) iff for every $\mathcal{S}$-structure $\mathfrak{A}$ and assignment s in $\mathfrak{A}$ if $\beta^{\mathfrak{A}}[s] \in \{T, t\}$, for every $\beta \in \Gamma$, then $\alpha^{\mathfrak{A}}[s] \in \{T, t\}$. Whenever α is a semantic consequence of Γ in this sense, this will be expressed by $\Gamma \vDash_{\mathbf{L1}} \alpha$ ($\Gamma \vDash_{\mathbf{L2}} \alpha$).

Enough for the semantics of **L1** and **L2**. In the next section, we will present their proof-theoretic counterparts, providing each logic with a corresponding sound and complete natural deduction system.

5. Two natural deduction systems

The systems $\mathcal{D}_{\mathbf{L1}}$ and $\mathcal{D}_{\mathbf{L2}}$ below correspond respectively to the logics **L1** and **L2** and result from adding the following rules governing the behavior of ! to an appropriate set of natural deduction rules for classical first-order logic (e.g., the ones presented in [26]):

System $\mathcal{D}_{\mathbf{L1}}$

$$\frac{}{!\mathbf{E}t}\ !\mathbf{E} \qquad \frac{!Pt_1 \dots t_n}{\mathbf{E}t_1 \wedge \cdots \wedge \mathbf{E}t_n}\ !\mathbf{E}1 \qquad \frac{}{!!\alpha}\ !!\mathrm{I} \qquad \frac{!\alpha}{!\neg\alpha}\ !\neg$$

$$\frac{!\alpha \qquad !\beta}{!(\alpha \wedge \beta)}\ ! \wedge \mathrm{I}1.1 \qquad \frac{\neg\alpha \qquad !\alpha}{!(\alpha \wedge \beta)}\ ! \wedge \mathrm{I}1.2 \qquad \frac{\neg\beta \qquad !\beta}{!(\alpha \wedge \beta)}\ ! \wedge \mathrm{I}1.3$$

$$\frac{!(\alpha \wedge \beta) \qquad \neg!\alpha}{\neg\beta \wedge !\beta}\ ! \wedge \mathrm{E}1.1 \qquad \frac{!(\alpha \wedge \beta) \qquad \neg!\beta}{\neg\alpha \wedge !\alpha}\ ! \wedge \mathrm{E}1.2$$

$$\frac{\forall x!\alpha}{!\forall x\alpha}\ !\forall\mathrm{I}1.1 \qquad \frac{\exists x(\neg\alpha \wedge !\alpha)}{!\forall x\alpha}\ !\forall\mathrm{I}1.2$$

$$\frac{!\forall x\alpha \qquad \exists x\neg!\alpha}{\exists x(\neg\alpha \wedge !\alpha)}\ !\forall\mathrm{E}1$$

[16] However, since the formal language we are working with also includes the existence predicate **E** among its logical primitives, by restricting $\forall$ and $\exists$ to **E**, it is possible to define a new pair of quantifiers $\forall_{\mathbf{E}}$ and $\exists_{\mathbf{E}}$ that behave as in dual domain free logics.

In $\mathbf{E}1$, P must be distinct from $\mathbf{E}$.[17]

SYSTEM $\mathcal{D}_{\mathbf{L2}}$

$$\frac{}{!\mathbf{E}t}\ !\mathbf{E} \qquad \dfrac{!Pt_1 \dots t_n}{\mathbf{E}t_1 \vee \cdots \vee \mathbf{E}t_n}\ !\mathbf{E}2 \qquad \frac{}{!!\alpha}\ !!\mathrm{I} \qquad \dfrac{!\alpha}{!\neg\alpha}\ !\neg$$

$$\frac{!\alpha \qquad !\beta}{!(\alpha \wedge \beta)}\ !\wedge\mathrm{I}2.1 \qquad \frac{\beta \qquad !\alpha}{!(\alpha \wedge \beta)}\ !\wedge\mathrm{I}2.2 \qquad \frac{\alpha \qquad !\beta}{!(\alpha \wedge \beta)}\ !\wedge\mathrm{I}2.3$$

$$\frac{!(\alpha \wedge \beta) \qquad \neg!\alpha}{\alpha \wedge !\beta}\ !\wedge\mathrm{E}2.1 \qquad \frac{!(\alpha \wedge \beta) \qquad \neg!\beta}{\beta \wedge !\alpha}\ !\wedge\mathrm{E}2.2$$

$$\frac{\forall x!\alpha}{!\forall x\alpha}\ !\forall\mathrm{I}2.1 \qquad \frac{\exists x!\alpha \qquad \forall x(\alpha \vee !\alpha)}{!\forall x\alpha}\ !\forall\mathrm{I}2.2$$

$$\frac{!\forall x\alpha \qquad \exists x\neg!\alpha}{\exists x!\alpha \wedge \forall x(\alpha \vee !\alpha)}\ !\forall\mathrm{E}2$$

As in $\mathcal{D}_{\mathbf{L1}}$, P must be distinct from $\mathbf{E}$ in rule $!\mathbf{E}2$.

Henceforth, we shall use the symbol $\mathbf{L}$ to refer to either $\mathbf{L1}$ or $\mathbf{L2}$, and write $\mathbf{L1}$, $\mathbf{L2}$, and $\mathbf{L}$ instead of respectively $\mathcal{D}_{\mathbf{L1}}$, $\mathcal{D}_{\mathbf{L2}}$, and $\mathcal{D}_{\mathbf{L}}$ (which, as in the case of $\mathbf{L}$, also refers indistinctively to either $\mathcal{D}_{\mathbf{L1}}$ or $\mathcal{D}_{\mathbf{L2}}$). The reader should encounter no difficulties to determine whether it is the logic or the corresponding deductive system that is being talked about in each context. As usual, the notion of a derivation in $\mathbf{L}$ can be inductively defined in terms of trees by adapting the definition in [26, p. 34]. We shall use the notation $\Gamma \vdash_{\mathbf{L}} \alpha$ to express that there is a derivation Θ in $\mathbf{L}$ such that α is the bottommost formula in Θ (its *conclusion*) and all of Θ's (undischarged) hypotheses belong to Γ.

Here is a couple interesting facts that hold for **L1** and **L2**:

PROPOSITION 5.1. *Let $\mathcal{S}$ be a first-order signature and $\Gamma \cup \{\alpha, \beta\} \subseteq Form(\mathcal{S})$. Then the following facts hold for* **L1** *and* **L2**:

1. *For every $\alpha \in Form(\mathcal{S})$, $\forall x\mathbf{E}x \vdash_{\mathbf{L}} !\alpha$.*
2. *For every $\alpha \in Form(\mathcal{S})$ such that* **E** *does occur in α, $\forall x\neg\mathbf{E}x \vdash_{\mathbf{L}} \neg!\alpha$.*
3. *Let* **CL** *denote the classical subsystem of* **L**. *If neither* **E** *nor* ! *occur in $\Gamma \cup \{\alpha\}$ and $\Gamma \vdash_{\mathbf{CL}} \alpha$, then*:
   a. *$\Gamma, \forall x\mathbf{E}x \vdash_{\mathbf{L}} \alpha \wedge !\alpha$; and*
   b. *$\Gamma, \forall x\neg\mathbf{E}x \vdash_{\mathbf{L}} \alpha \wedge \neg!\alpha$.*

---

[17] If $\mathbf{E}$ were allowed to be replaced by P in rule $!\mathbf{E}1$, then no term t could be such that $\neg\mathbf{E}t$ held (on pain of triviality). For since $\vdash_{\mathcal{D}_{\mathbf{L1}}} \forall x!\mathbf{E}x$ (by applying $\forall I$ to $!\mathbf{E}x$, which is an instance of $!\mathbf{E}$), $!\mathbf{E}1$ would then allow proving $\forall x\mathbf{E}x$.

Proof. *Ad* 1. It follows by induction on the complexity of α from the fact that $\forall x\mathbf{E}x \vdash_{\mathbf{L}} !Pt_1 \dots t_n$, for every term t, and the following !-propagation properties, which are immediate consequences of rules !**E**, !**E**1 (!**E**2), !$\neg$, and some of the !-introduction rules of **L1** and **L2**:

$$\vdash_{\mathbf{L}} !\mathbf{E}t \qquad !\alpha \vdash_{\mathbf{L}} !\neg\alpha \qquad !\alpha, !\beta \vdash_{\mathbf{L}} !(\alpha \wedge \beta) \qquad \forall x!\alpha \vdash_{\mathbf{L}} !\forall x\alpha$$

Ad 2. Likewise, it can be proven by induction on the complexity of α and:

$$\forall x\neg\mathbf{E}x \vdash_{\mathbf{L}} \neg!Pt_1 \dots t_n \qquad \neg!\alpha \vdash_{\mathbf{L}} \neg!\neg\alpha \qquad \neg!\alpha, \neg!\beta \vdash_{\mathbf{L}} \neg!(\alpha \wedge \beta)$$
$$\forall x\neg!\alpha \vdash_{\mathbf{L}} \neg!\forall x\alpha$$

which can be proven by rules !**E**, !**E**1 (!**E**2), !**E**$\neg$, and the !-elimination rules of **L1** and **L2**. □

As the reader might have noticed, both **L1** and **L2** are very close to classical logic. In fact, not only do they extend classical logic, but do so in a conservative manner:

Proposition 5.2. **L1** *and* **L2** *are conservative extensions of classical logic, i.e.:*

$$\Gamma \vdash_{\mathbf{CL}} \alpha \text{ iff } \Gamma \vdash_{\mathbf{L}} \alpha$$

for every $\Gamma \cup \{\alpha\}$ *such that neither* **E** *nor* ! *occur in* $\Gamma \cup \{\alpha\}$.

Proof. The left-to-right direction follows immediately from the fact **L** contains every rule of classical first-order logic. As for the right-to-left direction, we shall make use of the completeness of **L**, which will be proved in the next section (see Theorem 6.1): suppose that $\Gamma \nvdash_{\mathbf{CL}} \alpha$. By the completeness of classical logic, there is a classical structure $\mathfrak{A} = \langle A, \mathfrak{a}\rangle$ and an assignment s in $\mathfrak{A}$ such that $\mathfrak{A}, s \vDash \beta$, for every $\beta \in \Gamma$ and $\mathfrak{A}, s \nvDash \alpha$. Let $\mathfrak{A}_{\mathbf{L}} = \langle A, A_{\emptyset}, \mathfrak{a}\rangle$ be the **L**-structure that is just like $\mathfrak{A}$ except that $A_{\emptyset} = \emptyset$. It follows by induction on the complexity of γ that $\gamma^{\mathfrak{A}_{\mathbf{L}}}[s] = T$ iff $\mathfrak{A}, s \vDash \gamma$, and $\gamma^{\mathfrak{A}_{\mathbf{L}}}[s] = F$ iff $\mathfrak{A}, s \nvDash \gamma$, for every γ in which neither **E** nor ! occurs. Therefore, $\beta^{\mathfrak{A}_{\mathbf{L}}}[s] = T$, for every $\beta \in \Gamma$, and $\alpha^{\mathfrak{A}_{\mathbf{L}}}[s] = F$. By the completeness of **L**, we may conclude that $\Gamma \nvdash_{\mathbf{L}} \alpha$. □

Given the semantic conditions of !, it should be clear that usual the formulation of the replacement theorem, i.e., $\alpha \leftrightarrow \beta \vDash_{\mathbf{L}} \gamma \leftrightarrow \gamma[\beta\|\alpha]$, does not hold in general, as $!\alpha \leftrightarrow !\beta$ will receive the value F when $\alpha^{\mathfrak{A}}[s] = T$ $(= t)$ and $\beta^{\mathfrak{A}}[s] = t$ $(= T)$ or $\alpha^{\mathfrak{A}}[s] = F$ $(= f)$ or $\beta^{\mathfrak{A}}[s] = f$ $(= F)$ (and in both cases $(\alpha \leftrightarrow \beta)^{\mathfrak{A}}[s] \in \{T, t\}$). However, the very presence of ! allows us to define a stronger equivalence relation that, unlike $\to$, permits replacing equivalents formulas for one another in every sentential context:

$$\alpha \equiv_{!} \beta \;:=\; (\alpha \leftrightarrow \beta) \wedge (!\alpha \leftrightarrow !\beta)$$

By induction on the complexity of γ we obtain:

Proposition 5.3. *Let $\gamma[\beta\|\alpha]$ denote a formula that results from replacing zero or more occurrences of α in γ by β. If none of the variables free α or β is also free in γ, then $\alpha \equiv_! \beta \vdash_{\mathbf{L}} \gamma \equiv_! \gamma[\beta\|\alpha]$.*

Remark 5.1. It is worth noticing at this point that **L1** and **L2** are not the only logics that comprise our framework. For instance, in addition to the truth value orders of **L1** and **L2**, there at least two alternatives, namely:

factually false $<$ *fictionally false* $<$ *factually true* $<$ *fictionally true*
fictionally false $<$ *factually false* $<$ *fictionally true* $<$ *factually true*

Had we also taken these two orders into account, we would end up with not two, but at least four distinct logics — and the number would have increased significantly were we to formulate two logics for each order according to the whether the atomic formulas are interpreted according to (**P1**) or (**P2**).[18]

Investigating each logic in our framework would certainly be both instructive and interesting, but due to the lack of space we have decided to focus on just two.[19]

Nonetheless, in view of the results in this paper, we believe that the task of explicitly formulating other logics in our framework by adapting what has been done for **L1** and **L2** would pose no serious difficulties. So, if one comes up with specific reasons that favor, say, combining (**P2**) with **L1**'s semantic clauses for $\wedge$ and $\forall$, then the task of formulating the corresponding deductive system will be almost straightforward.

6. Soundness and completeness

Let us take up now the task of proving the soundness and the completeness of the two systems above. Since several of the auxiliary results we shall make use of carry over from classical logic to **L1** and **L2** (with some small adjustments), we shall not present their proofs in full detail here. In most cases, those results will be merely enunciated for the sake of reference within other proofs.

The proofs of the soundness of **L1** and **L2** with respect to the corresponding semantics follow the standard procedure of showing that all rules preserve the designated values (viz., T and t). As with classical logic, showing this to be case for the quantifier rules requires the following lemma.

[18] Notice, for instance, that nothing prevents one from interpreting atomic formulas according to (**P2**) while interpreting $\wedge$ and $\forall$ as in **L1**.

[19] Our specific choice was inspired by Ivlev [11] and Kearns' [13] approaches to four-valued modal logic, in which the values are ordered as follows:

impossible $<$ *contingently false* $<$ *contingently true* $<$ *necessarily true*

Lemma 6.1. *Let $\mathfrak{A}$ be an $\mathcal{S}$-structure and s an assignment in $\mathfrak{A}$. Let t belong to Term($\mathcal{S}$), and define s' to be the assignment $s(\overline{s}(t)/x)$ (that is, the assignment that is just like s except that it assigns $\overline{s}(t)$ to the variable x). If t is free for x in $\alpha \in$ Form($\mathcal{S}$), then $\alpha^{\mathfrak{A}}[s'] = (\alpha[t/x])^{\mathfrak{A}}[s]$.*

The proofs of the completeness of **L1** and **L2** are also quite similar to the proof of the completeness of classical logic, and so we shall focus on those parts where the main differences occur. The following two definitions might be familiar to the reader:

Definition 6.1. Let $\mathcal{S}$ be a first-order signature and $\Delta \subseteq$ *Form*($\mathcal{S}$). Δ is a *maximally nontrivial theory* of **L** if: (i) Δ is nontrivial, i.e., $\Delta \nvdash_{\mathbf{L}} \alpha$, for some $\alpha \in$ *Form*($\mathcal{S}$); and (ii) for every $\alpha \in$ *Form*($\mathcal{S}$), if $\alpha \notin \Delta$, then $\Delta \cup \{\alpha\}$ is trivial.

Definition 6.2. Let $\mathcal{S} = \langle \mathcal{C}, \mathcal{F}, \mathcal{P} \rangle$ be a first-order signature and suppose that $\Delta \subseteq$ *Form*($\mathcal{S}$). Then, Δ is called a *Henkin set* if for every formula α and variable x, $\Delta \vdash_{\mathbf{L}} \forall x \alpha$ iff $\Delta \vdash_{\mathbf{L}} \alpha[c/x]$, for every $c \in \mathcal{C}$.

Notice that every maximally non-trivial theory Δ is indeed a *theory* in the sense that if $\Delta \vdash_{\mathbf{L}} \alpha$, then $\alpha \in \Delta$. Notice further that Δ is also a *prime* set, for either $\alpha \in \Delta$ or $\beta \in \Delta$ whenever $\alpha \vee \beta \in \Delta$ (and, in particular, $\alpha \in \Delta$ or $\neg\alpha \in \Delta$, for every α). Because Δ is non-trivial, $\neg\alpha \in \Delta$ iff $\alpha \notin \Delta$. Moreover, if Δ is a Henkin set, it satisfies the following property, which guarantees the existence of a witness c for every existential formula that belongs to Δ: $\exists x \alpha \in \Delta$ iff $\alpha[c/x] \in \Delta$, for some constant c. Below we shall make use of these properties without explicitly mentioning them.

The following result ensures that every non-trivial set can be extended to a maximally non-trivial Henkin theory. Because its proof has no significant differences from the proof of the corresponding classical result, we shall not bother presenting it here:

Proposition 6.1. *Let $\mathcal{S} = \langle \mathcal{C}, \mathcal{F}, \mathcal{P} \rangle$ be a first-order signature and suppose that $\alpha \in$ Form($\mathcal{S}$). Let $\Gamma \subseteq$ Form($\mathcal{S}$) be nontrivial. Then, there is a first-order signature $\mathcal{S}^+ = \langle \mathcal{C}^+, \mathcal{F}, \mathcal{P} \rangle$ and a set $\Delta \subseteq$ Form($\mathcal{S}^+$) such that $\mathcal{C} \subseteq \mathcal{C}^+$, $\Gamma \subseteq \Delta$, and Δ is a maximally nontrivial Henkin theory.*

We shall now proceed to the last step of the proof of the completeness of **L1** and **L2**, which consists in showing that every maximally non-trivial Henkin theory has a *canonical model*. Before we do so, however, it is worth noticing that whenever Δ is a maximally non-trivial Henkin theory, we can define an equivalence relation $\sim$ between terms in the following way:

$$t \sim u \text{ iff } \Delta \vdash_{\mathbf{L}} t = u$$

That $\sim$ is an equivalence relation follows immediately from the natural deduction rule for $=$. Those very same rules also ensure that: if $t_1 \sim t'_1, \ldots, t_n \sim t'_n$, then $ft_1 \dots t_n = ft'_1 \dots t'_n \in \Delta$, for every function letter f; and if $t_1 \sim t'_1, \dots, t_n \sim t'_n$, then $Pt_1 \dots t_n \in \Delta$ iff $Pt'_1 \dots t'_n \in \Delta$, for every predicate letter P.

DEFINITION 6.3. Let $\mathcal{S} = \langle \mathcal{C}, \mathcal{F}, \mathcal{P} \rangle$ be a first-order signature, and suppose that $\Delta \subseteq Form(\mathcal{S})$ is a maximally nontrivial Henkin theory. For each term $t \in Term(\mathcal{S})$, let $[t]$ be the equivalence class $\{u \in Term(\mathcal{S}) : t \sim u\}$. Then, the *canonical structure* $\mathfrak{A} = \langle A, A_\emptyset, ^{\mathfrak{a}} \rangle$ of Δ is defined by:

- $A = \{[t] : \mathbf{E}t \in \Delta\}$ and $A_\emptyset = \{[t] : \mathbf{E}t \notin \Delta\}$;
- For each $c \in \mathcal{C}$, $c^{\mathfrak{A}} = [c]$;
- For each $f \in \mathcal{F}_n$, $f^{\mathfrak{A}}([t_1], \dots, [t_n]) = [ft_1 \dots t_n]$;
- For each $P \in \mathcal{P}_n$, $P^{\mathfrak{A}} = \{\langle [t_1], \dots, [t_n] \rangle : Pt_1 \dots t_n \in \Delta\}$.

Given a maximally nontrivial Henkin theory Δ and its canonical structure $\mathfrak{A}$, we shall call the assignment s such that $s(x) = [x]$, for every variable x, *the canonical assignment* of $\mathfrak{A}$. It is straightforward to prove by induction on the complexity of t that $\overline{s}(t) = [t]$, for every $t \in Term(\mathcal{S})$. Notice that since Δ is consistent, it follows that $[t] \in A$ iff $[t] \notin A_\emptyset$, for every term t. This is an immediate consequence of Definition 6.3 above, which guarantees that $A \cap A_\emptyset = \emptyset$. Notice further that since $\mathbf{E}t \in \Delta$ or $\mathbf{E}t \notin \Delta$, for every t, $A \cup A_\emptyset = \{[t] : t \in Term(\mathcal{S})\} = |\mathfrak{A}|$. Moreover, due to the way $\sim$ was defined, $\langle [t], [u] \rangle \in =^{\mathfrak{A}} [s]$ iff $t = u \in \Delta$ iff $[t] = [u]$. Hence:

$$=^{\mathfrak{A}} [s] = \{\langle [t], [t] \rangle : t \in Term(\mathcal{S})\}$$

Finally, the properties of $\sim$ mentioned above guarantee that the definitions of $P^{\mathfrak{A}}$ and $f^{\mathfrak{A}}$ do not depend on the representatives $[t_1], \dots, [t_n]$. As a result, if Δ is a maximally nontrivial Henkin theory, its canonical structure $\mathfrak{A}$ satisfies the conditions of Definition 4.1.

Next, we establish the very last result before we can conclude the proof of the completeness of **L1** and **L2**. Before proceeding to that task, though, it is worth mentioning a couple of results concerning *alphabetic variants* that we will be implicitly used throughout the proof. These two facts assert respectively the semantic and proof-theoretic equivalence of formulas that differ only in some of their bound variables (e.g., $\forall x(Px \wedge \exists z Qzx)$ and $\forall y(Py \wedge \exists w Qwy)$):

- If β is an alphabetic variant of α, then $\alpha^{\mathfrak{A}}[s] = \beta^{\mathfrak{A}}[s]$; and
- If β is an alphabetic variant of α, then $\alpha \vdash_{\mathbf{L}} \beta$ and $\beta \vdash_{\mathbf{L}} \alpha$.

PROPOSITION 6.2. *Let $\mathcal{S}$ be a first-order signature and suppose that $\Delta \subseteq Form(\mathcal{S})$ is a maximally nontrivial Henkin theory. Let $\mathfrak{A}$ be the canonical structure of Δ, and s its canonical assignment. Then the following facts hold for every $\alpha \in Form(\mathcal{S})$:*

1. *$\alpha^{\mathfrak{A}}[s] = T$ iff $\alpha, !\alpha \in \Delta$;*
2. *$\alpha^{\mathfrak{A}}[s] = t$ iff $\alpha, \neg!\alpha \in \Delta$;*

3. $\alpha^{\mathfrak{A}}[s] = f$ *iff* $\neg\alpha, \neg!\alpha \in \Delta$;
4. $\alpha^{\mathfrak{A}}[s] = F$ *iff* $\neg\alpha, !\alpha \in \Delta$.

Proof. The result follows by induction on the length of α, but the proof varies at certain points according to whether $\boldsymbol{L} = \mathbf{L1}$ or $\boldsymbol{L} = \mathbf{L2}$. However, we shall only present here the proof for $\boldsymbol{L} = \mathbf{L1}$, which can be easily adapted to $\mathbf{L2}$.

Let α be $Pt_1 \dots t_n$, where P is distinct from $\mathbf{E}$. *Ad* 1. $\alpha^{\mathfrak{A}}[s] = T$ iff $\langle [t_1], \dots, [t_n] \rangle \in P^{\mathfrak{A}}$ and $[t_i] \in A$, for every $1 \leqslant i \leqslant n$, iff $Pt_1 \dots t_n \in \Delta$ and $\mathbf{E}t_i \in \Delta$, for every $1 \leqslant i \leqslant n$, iff $Pt_1 \dots t_n \in \Delta$ and $\mathbf{E}t_1 \wedge \cdots \wedge \mathbf{E}t_n \in \Delta$ iff $Pt_1 \dots t_n \in \Delta$ and $!Pt_1 \dots t_n \in \Delta$. Where the last step is justified by rule $!\mathbf{E}1$.

Ad 2. $\alpha^{\mathfrak{A}}[s] = t$ iff $\langle [t_1], \dots, [t_n] \rangle \in P^{\mathfrak{A}}$ and $[t_i] \notin \mathbf{E}^{\mathfrak{A}}$, for some $1 \leqslant i \leqslant n$, iff $Pt_1 \dots t_n \in \Delta$ and $\neg\mathbf{E}t_i \in \Delta$, for some $1 \leqslant i \leqslant n$, iff $Pt_1 \dots t_n \in \Delta$ and $\neg!Pt_1 \dots t_n \in \Delta$. Where the last step is justified by rule $!\mathbf{E}1$.

Ad 3. $\alpha^{\mathfrak{A}}[s] = f$ iff $\langle [t_1], \dots, [t_n] \rangle \notin P^{\mathfrak{A}}$ and $[t_i] \notin \mathbf{E}^{\mathfrak{A}}$, for some $1 \leqslant i \leqslant n$, iff $\neg Pt_1 \dots t_n \in \Delta$ and $\neg\mathbf{E}t_i \in \Delta$, for some $1 \leqslant i \leqslant n$, iff $\neg Pt_1 \dots t_n \in \Delta$ and $\neg!Pt_1 \dots t_n \in \Delta$. Where the last step is justified by rule $!\mathbf{E}1$.

Ad 4. $\alpha^{\mathfrak{A}}[s] = F$ iff $\langle [t_1], \dots, [t_n] \rangle \notin P^{\mathfrak{A}}$ and $[t_i] \in \mathbf{E}^{\mathfrak{A}}$, for every $1 \leqslant i \leqslant n$, iff $\neg Pt_1 \dots t_n \in \Delta$ and $\mathbf{E}t_i \in \Delta$, for every $1 \leqslant i \leqslant n$, iff $\neg Pt_1 \dots t_n \in \Delta$ and $\mathbf{E}t_1 \wedge \cdots \wedge \mathbf{E}t_n \in \Delta$ iff $\neg Pt_1 \dots t_n \in \Delta$ and $!Pt_1 \dots t_n \in \Delta$. Where the last step is justified by rule $!\mathbf{E}1$.

Let α be $\mathbf{E}t$. *Ad* 1. $\alpha^{\mathfrak{A}}[s] = T$ iff $[t] \in A$ iff $\mathbf{E}t \in \Delta$ iff $\mathbf{E}t, !\mathbf{E}t \in \Delta$. Where the last step is justified by rule $!\mathbf{E}1$.

Ad 2. $\alpha^{\mathfrak{A}}[s] = F$ iff $[t] \in A_\emptyset$ iff $\mathbf{E}t \notin \Delta$ iff $\neg\mathbf{E}t \in \Delta$ iff $\neg\mathbf{E}t, !\mathbf{E}t \in \Delta$. Where the last step is justified by rule $!\mathbf{E}$.

Let α be $\neg\beta$. *Ad* 1. $\alpha^{\mathfrak{A}}[s] = T$ iff $\beta^{\mathfrak{A}}[s] = F$ iff $\neg\beta, !\beta \in \Delta$ (by IH) iff $\neg\beta, !\neg\beta$, i.e., $\alpha, !\alpha \in \Delta$. Where the last step is justified by rule $!\neg$.

Ad 2. $\alpha^{\mathfrak{A}}[s] = t$ iff $\beta^{\mathfrak{A}}[s] = f$ iff $\neg\beta, \neg!\beta \in \Delta$ (by IH) iff $\neg\beta, \neg!\neg\beta \in \Delta$, i.e., $\alpha, \neg!\alpha \in \Delta$. Where the last step is justified by rule $!\neg$.

Ad 3. $\alpha^{\mathfrak{A}}[s] = f$ iff $\beta^{\mathfrak{A}}[s] = t$ iff $\beta, \neg!\beta \in \Delta$ (by IH) iff $\neg\neg\beta, \neg!\neg\beta \in \Delta$, i.e., $\neg\alpha, \neg!\alpha \in \Delta$. Where the last step is justified by rule $!\neg$.

Ad 4. $\alpha^{\mathfrak{A}}[s] = F$ iff $\beta^{\mathfrak{A}}[s] = T$ iff $\beta, !\beta \in \Delta$ (by IH) iff $\neg\neg\beta, !\neg\beta \in \Delta$, i.e., $\neg\alpha, !\alpha \in \Delta$. Where the last step is justified by rule $!\neg$.

Let α be $!\beta$. *Ad* 1. $\alpha^{\mathfrak{A}}[s] = T$ iff $\beta^{\mathfrak{A}}[s] = T$ or $\beta^{\mathfrak{A}}[s] = F$ iff $\beta, !\beta \in \Delta$ or $\neg\beta, !\beta \in \Delta$ (by IH) iff $!\beta, !!\beta \in \Delta$, i.e., $\alpha, !\alpha \in \Delta$. Where the last step is justified by rule $!!$I.

Ad 2. $\alpha^{\mathfrak{A}}[s] = F$ iff $\beta^{\mathfrak{A}}[s] = t$ or $\beta^{\mathfrak{A}}[s] = f$ iff $\beta, \neg!\beta \in \Delta$ or $\neg\beta, \neg!\beta \in \Delta$ (by IH) iff $\neg!\beta, !!\beta \in \Delta$. Where the last step is justified by rule $!!$I.

Let α be $\beta \wedge \gamma$. *Ad* 1. $\alpha^{\mathfrak{A}}[s] = T$ iff $\beta^{\mathfrak{A}}[s] = T$ and $\gamma^{\mathfrak{A}}[s] = T$ iff $\beta, !\beta \in \Delta$ and $\gamma, !\gamma \in \Delta$ (by IH) iff $\beta \wedge \gamma, !(\beta \wedge \gamma) \in \Delta$, i.e., $\alpha, !\alpha \in \Delta$. Where the left-to-right direction of the last step is justified by rule $!\wedge$I1.1, and its right-to-left direction by rules $!\wedge$E1.1 and $!\wedge$E1.2.

Ad 2. $\alpha^{\mathfrak{A}}[s] = t$ iff $\beta^{\mathfrak{A}}[s], \gamma^{\mathfrak{A}}[s] \in \{T, t\}$ and either $\beta^{\mathfrak{A}}[s] = t$ or $\gamma^{\mathfrak{A}}[s] = t$ iff $\beta, \gamma \in \Delta$ and either $\neg!\beta \in \Delta$ or $\neg!\gamma \in \Delta$ (by IH) iff $\beta \wedge \gamma, \neg!(\beta \wedge \gamma) \in \Delta$, i.e.,

$\alpha \in \Delta, \neg!\alpha \in \Delta$. Where the left-to-right direction of the last step is justified by rules $!\wedge$E1.1 and $!\wedge$E1.2, and its right-to-left direction by rule $!\wedge$I1.1.

Ad 3. $\alpha^{\mathfrak{A}}[s] = f$ iff $\beta^{\mathfrak{A}}[s], \gamma^{\mathfrak{A}}[s] \notin \{F\}$ and either $\beta^{\mathfrak{A}}[s] = f$ or $\gamma^{\mathfrak{A}}[s] = f$ iff $(\neg\beta \notin \Delta$ or $!\beta \notin \Delta)$, $(\neg\gamma \notin \Delta$ or $!\gamma \notin \Delta)$, and either $\neg\beta \in \Delta$ or $\neg\gamma \in \Delta$ (by IH) iff $\neg(\beta \wedge \gamma), \neg!(\beta \wedge \gamma)$, i.e., $\neg\alpha, \neg!\alpha \in \Delta$. Where the left-to-right direction of the last step is justified by rules $!\wedge$E1.1 and $!\wedge$E1.2, and its right-to-left direction by rules $!\wedge$I1.2, and $!\wedge$I1.3.

Ad 4. $\alpha^{\mathfrak{A}}[s] = F$ iff $\beta^{\mathfrak{A}}[s] = F$ or $\gamma^{\mathfrak{A}}[s] = F$ iff $\neg\beta, !\beta \in \Delta$ or $\neg\gamma, !\gamma \in \Delta$ iff $\neg(\beta \wedge \gamma), !(\beta \wedge \gamma) \in \Delta$, i.e., $\neg\alpha, !\alpha \in \Delta$. Where the left-to-right direction of the last step is justified by rules $!\wedge$I1.2 and $!\wedge$I1.3, and its right-to-left direction by rules $!\wedge$E1.1 and $!\wedge$E1.2.

Let α be $\forall x\beta$. *Ad* 1. "$\Rightarrow$" If $\alpha^{\mathfrak{A}}[s] = T$, then $\beta^{\mathfrak{A}}[s_x^{[t]}] = T$, for every $t \in Term(\mathcal{S})$. In particular, $\beta^{\mathfrak{A}}[s_x^{[c]}] = T$, for every $c \in mathcalC$. By Lemma 6.1, $(\beta[c/x])^{\mathfrak{A}}[s] = T$, for every c. Hence, $\beta[c/x], !\beta[c/x] \in \Delta$, for every c (by IH). Since Δ is a Henkin set, it follows that $\forall x\beta, \forall x!\beta \in \Delta$. By rule $!\forall$I1.1, we then have $\forall x\beta, !\forall x\beta \in \Delta$. "($\Leftarrow$" If $\forall x\beta, !\forall x\beta \in \Delta$, then $\forall x\beta, \forall x!\beta \in \Delta$ (by rule $!\forall$E1). Let $t \in Term(\mathcal{S})$ and let β^* be an alphabetic variant of β such that t is free for x in β^*. Thus, $\beta^*[t/x], !\beta^*[t/x] \in \Delta$. By IH, it follows that $(\beta^*[t/x])^{\mathfrak{A}}[s] = T$, and so $\beta^{*\mathfrak{A}}[s_x^{[t]}] = T$ (by Lemma 6.1). Hence, $\beta^{\mathfrak{A}}[s_x^{[t]}] = T$, and since t was arbitrary, $\beta^{\mathfrak{A}}[s_x^{[t]}] = T$, for every t; that is, $(\forall x\beta)^{\mathfrak{A}}[s] = T$.

Ad 2. "$\Rightarrow$" If $\alpha^{\mathfrak{A}}[s] = t$, then $\beta^{\mathfrak{A}}[s_x^{[t]}] \in \{T, t\}$, for every $t \in Term(\mathcal{S})$, and $\beta^{\mathfrak{A}}[s_x^{[t']}] = t$, for some $t' \in Term(\mathcal{S})$. In particular, $\beta^{\mathfrak{A}}[s_x^{[c]}] \in \{T, t\}$, for every $c \in \mathcal{C}$. By Lemma 6.1, it follows that $(\beta[c/x])^{\mathfrak{A}}[s] \in \{T, t\}$, for every c, and $(\beta^*[t'/x])^{\mathfrak{A}}[s] = t$, where β^* is an alphabetic variant of β such that t' is free for x in β^*. By IH, $\beta[c/x] \in \Delta$, for every c, and $\beta^*[t'/x], \neg!\beta^*[t'/x] \in \Delta$. Since Δ is a Henkin set, it follows that $\forall x\beta \in \Delta$. And since $\neg!\beta^*[t'/x] \in \Delta$, $\exists x\neg!\beta^* \in \Delta$, and so $\exists x\neg!\beta \in \Delta$. Therefore, $\forall x\beta, \exists x\neg!\beta \in \Delta$. By rule $!\forall$E1, $\neg!\forall x\beta \in \Delta$ (for otherwise $\exists x\neg\alpha \in \Delta$), and so $\forall x\beta, \neg!\forall x\beta \in \Delta$. "$\Leftarrow$" If $\forall x\beta, \neg!\forall x\beta \in \Delta$, then $\neg\forall x!\beta \in \Delta$ (by rule $!\forall$I1.1). Thus, $\exists x\neg!\beta \in \Delta$. Let $c \in \mathcal{C}$ be such that $\neg!\beta[c/x] \in \Delta$ (since Δ is a Henkin set). Since $\beta[c/x] \in \Delta$, it follows by IH that $(\beta[c/x])^{\mathfrak{A}}[s] = t$, and so $\beta^{\mathfrak{A}}[s_x^{[c]}] = t$ (by Lemma 6.1). Now, recall that $\forall x\beta \in \Delta$ and let $t \in Term(\mathcal{S})$. Let β^* be an alphabetic variant of β such that t is free for x in β^*. Hence, $\beta^*[t/x] \in \Delta$, and so $(\beta^*[t/x])^{\mathfrak{A}}[s] \in \{T, t\}$. By Lemma 6.1, $\beta^{*\mathfrak{A}}[s_x^{[t]}] \in \{T, t\}$, and therefore $\beta^{\mathfrak{A}}[s_x^{[t]}] \in \{T, t\}$. Since t was arbitrary, $\beta^{\mathfrak{A}}[s_x^{[t]}] \in \{T, t\}$, for every t. Therefore, $(\forall x\beta)^{\mathfrak{A}}[s] = t$.

Ad 3. "$\Rightarrow$" If $\alpha^{\mathfrak{A}}[s] = f$, then $\beta^{\mathfrak{A}}[s_x^{[t]}] \neq F$, for every $t \in Term(\mathcal{S})$, and $\beta^{\mathfrak{A}}[s_x^{[t']}] = f$, for some $t' \in Term(\mathcal{S})$. In particular, $\beta^{\mathfrak{A}}[s_x^{[c]}] \neq F$, for every $c \in \mathcal{C}$. By Lemma 6.1, it follows that $(\beta[c/x])^{\mathfrak{A}}[s] \neq F$, for every c, and $(\beta^*[t'/x])^{\mathfrak{A}}[s] = f$, where β^* is an alphabetic variant of β such that t' is free for x in β^*. By IH, $\neg\beta[c/x] \notin \Delta$ or $!\beta[c/x] \notin \Delta$, for every c, and so $\beta[c/x] \in \Delta$ or $\neg!\beta[c/x] \in \Delta$, for every c. Therefore, $\beta[c/x] \vee \neg!\beta[c/x] \in \Delta$, for every c, and since Δ is a Henkin set, $\forall x(\beta \vee \neg!\beta) \in \Delta$. Now, it follows from $(\beta^*[t'/x])^{\mathfrak{A}} = f$ that $\neg\beta^*[t'/x], \neg!\beta^*[t'/x] \in \Delta$ (also by IH). Thus, $\exists x\neg\beta^* \in \Delta$, and so $\exists x\neg\beta \in \Delta$. By rule !∀E1, we then have $\neg\forall x\beta, \neg!\forall x\beta \in \Delta$. "$\Leftarrow$" If $\neg\forall x\beta, \neg!\forall x\beta \in \Delta$, then $\forall x(\beta\vee\neg!\beta) \in \Delta$ (by rule !∀I1.2). Let $c \in \mathcal{C}$ be such that $\neg\beta[c/x] \in \Delta$ (since $\exists x\neg\beta \in \Delta$ and Δ is a Henkin set). Since $\forall x(\beta\vee\neg!\beta) \in \Delta$, it follows that $\neg\beta[c/x], \neg!\beta[c/x] \in \Delta$. By IH, $(\beta[c/x])^{\mathfrak{A}}[s] = f$, and so $\beta^{\mathfrak{A}}[s_x^{[c]}] = f$ (by Lemma 6.1). Now, let $t \in \mathit{Term}(\mathcal{S})$ and let β^* be an alphabetic variant of β such that t is free for x in β^*. Since $\forall x(\beta^* \vee \neg!\beta^*) \in \Delta$, it follows that $\beta^*[t/x]\vee!\beta^*[t/x] \in \Delta$. Therefore, $\beta^*[t/x] \in \Delta$ or $\neg!\beta^*[t/x] \in \Delta$. By IH, $(\beta^*[t/x])^{\mathfrak{A}}[s] \neq F$, and so $\beta^{*\mathfrak{A}}[s_x^{[t]}] \neq F$ (by Lemma 6.1). Thus, $\beta^{\mathfrak{A}}[s_x^{[t]}] \neq F$, and since t was arbitrary, $\beta^{\mathfrak{A}}[s_x^{[t]}] \neq F$, for every t. Therefore, $(\forall x\beta)^{\mathfrak{A}}[s] = f$.

Ad 4. "$\Rightarrow$" If $\alpha^{\mathfrak{A}}[s] = F$, then $\beta^{\mathfrak{A}}[s_x^{[t]}] = F$, for some $t \in \mathit{Term}(\mathcal{S})$. Let β^* be an alphabetic variant of β such that t is free for x in β^*. By Lemma 6.1, $(\beta^*[t/x])^{\mathfrak{A}}[s] = F$. By IH, it then follows that $\neg\beta^*[t/x], !\beta^*[t/x] \in \Delta$. Thus, $\exists x(\neg\beta^* \wedge !\beta^*) \in \Delta$, and so $\exists x(\neg\beta \wedge !\beta) \in \Delta$. By rule !∀I1.2, we have $!\forall x\beta \in \Delta$. Therefore, $\neg\forall x\beta, !\forall x\beta \in \Delta$. "$\Leftarrow$" If $\neg\forall x\beta, !\forall x\beta \in \Delta$, then $\exists x(\neg\beta \wedge !\beta) \in \Delta$ (by assuming $\neg\exists x(\neg\beta \wedge !\beta) \in \Delta$ and applying rule !∀E1). Since Δ is a Henkin set, there is $c \in \mathcal{C}$ such that $\neg\beta[c/x] \wedge !\beta[c/x] \in \Delta$. By IH, $(\beta[c/x])^{\mathfrak{A}}[s] = F$, and so L $\beta^{\mathfrak{A}}[s_x^{[c]}] = F$ (by Lemma 6.1). Therefore, $(\forall x\beta)^{\mathfrak{A}}[s] = F$. □

We now have everything we need in order to prove the completeness of **L1** and **L2**:

Theorem 6.1. *Let $\mathcal{S} = \langle \mathcal{C}, \mathcal{F}, \mathcal{P}\rangle$ be a first-order signature and $\Gamma \cup \{\alpha\}$ be a subset of $\mathit{Form}(\mathcal{S})$. If $\Gamma \vDash_{\mathbf{L}} \alpha$, then $\Gamma \vdash_{\mathbf{L}} \alpha$.*

Proof. Suppose that $\Gamma \nvdash_{\mathbf{L}} \alpha$. Hence, $\Gamma \cup \{\neg\alpha\}$ is nontrivial. By Proposition 6.1, there is a first-order signature $\mathcal{S}^+ = \langle \mathcal{C}^+, \mathcal{F}, \mathcal{P}\rangle$ such that for some $\Delta \subseteq \mathit{Form}(\mathcal{S}^+)$ we have $\mathcal{C} \subseteq \mathcal{C}^+$, $\Gamma \cup \{\neg\alpha\} \subseteq \Delta$, and Δ is a maximally nontrivial Henkin theory. By Proposition 6.2, there is an $\mathcal{S}^+$-structure $\mathfrak{A}$ and an assignment s such that $\beta^{\mathfrak{A}}[s] = T$ or $\beta^{\mathfrak{A}}[s] = t$, for every $\beta \in \Delta$. In particular, $\neg\alpha^{\mathfrak{A}}[s] = T$ or $\neg\alpha^{\mathfrak{A}}[s] = t$. Thus, $\mathfrak{A}, s \vDash_{\mathbf{L}} \beta$, for every $\beta \in \Delta$, and $\mathfrak{A}, s \nvDash_{\mathbf{L}} \alpha$. Consider now the $\mathcal{S}$-reduct $\mathfrak{B}$ of $\mathfrak{A}$. Clearly, $\mathfrak{B}, s \vDash_{\mathbf{L}} \beta$, for every $\beta \in \Gamma$, while $\mathfrak{B}, s \nvDash_{\mathbf{L}} \alpha$. That is, $\Gamma \nvDash_{\mathbf{L}} \alpha$. □

7. Final remarks

In this paper we have proposed a semantic framework in which fictional statements are evaluated as either factually true (false) or fictionally true (false). The proposal was meant to explicitly address the intuition that declaring 'Socrates is a man' to be true does not come down to the same thing as declaring 'Sherlock

Holmes is a man' to be so. It also acknowledges the fact that we may and sometimes do reason about both real and fictional individuals in the same context, and so we presented two alternative ways to interpret mixed statements. As for the technical implementation of these ideas, we developed two different logics, **L1** and **L2**, that expand classical logic by including two new logical symbols, the operator ! and existence predicate **E**. The logics were provided with natural deduction systems which were then proved to be sound and complete.

Though we have presented two alternative logics, the problem of choosing between them cannot be easily solved. For consider the following sentence:

$$\text{Dom Quixote admires Carlos Magnus} \tag{16}$$

(14) is intuitively factually true because it is a sentence in which the subject of the action is a real individual while the object is a fictional one. Sentence (16), on the other hand, appears to be fictionally true because the subject of the action is fictional whereas its object is real. But our approach either forces both (14) and (16) to have factual values or both sentences to have fictional values.[20]

One way to produce a single more expressive logic would be by adopting a non-deterministic semantics, which have been extensively studied by, for example, [1]. In the case of complex formulas, the technique is as follows: if we don't know, for instance, whether (14) is fictionally true or factually true, we wouldn't need to assign it a single truth value; instead, (14) would get assigned a set of truth values whose elements are T and t.

But this approach faces at least two challenges. The first one is that it ends up expressing more differences in truth values than it would be desirable. For consider:

$$\exists y(Asy \wedge Cy) \tag{5f2}$$

Both (5f1) and (5f2) could eventually receive different truth values in a non-deterministic semantics. Of course, this drawback could be technically avoided by adding an axiom that forces formulas like (5f1) and (5f2) to be logically equivalent, which is, however, clearly *ad hoc*.

The second challenge is this: let b, c, d, and m be constants that stand for Tony Blair, Coriolanus, Dom Quixote and Carlos Magnus, respectively; and let A be the formal rendering of 'x admires y'. In a fictional interpretation of (14) and (16) we would then have:

$$Abc \wedge \neg !Abc \tag{14f}$$

$$Adm \wedge \neg !Adm \tag{16f}$$

Non-deterministic semantics allow assigning sets of truth values only to complex formulas. Thus, the non-deterministic aspect of (14f) and (16f) would be due to $\wedge$ and !, rather than to the predicate A or the individual constants. However, we take it to be clear that the difference in the truth values of (14) and (16)

[20] Again, we would like to thank the anonymous referee for this observation.

depends on whether the subject or the object of the binary relation 'x admires y' is a fictional character. But this is not possible in a non-deterministic semantics.

Another way to account for the difference in the truth values of (14f) and (16f) within the same logic would be to incorporate two versions of the ! operator. But again, if c and d have fictional references while b and m refer to real objects, we wouldn't be able to say that Abc is factually true and Adm is fictionally true. This means that in some usages collapsing occurrences of the ! operator might not be desirable. If we accept the *ad hoc* solutions of adopting non-deterministic semantics to force sentences like (5f1) and (5f2) to be logically equivalent, we could build a whole hierarchy of systems that aim to control the interactions of the ! operator similarly to what happens with the control of occurrences of modal operators in relational semantics. This looks like a promising line of future research.[21]

Now, consider:

Ana Karenina wore yellow stockings at her first meeting with Vronsky (17)

Tolstoy's novel *Ana Karenina* says nothing about the color of Ana Karenina's underwear at her first date with Vronsky, and there seems to be nothing beyond what is told in the novel itself that would help us decide the matter. Thus, it is unrealistic to require (17) to be either fictionally true or fictionally false, even though (17) turns out to be neither factually true nor factually false. Hence, our framework is not entirely satisfactory in that it does not allow for fictional *truth value gaps*.

Not only are most fictional stories incomplete, but some few pieces of fiction are inconsistent as well, describing some of their characters or the events that take place in their stories in contradictory ways (see, for instance, the stories in [7, 21]). Hence, if we are willing to allow for fictional truth value gaps, there seems to be no reason not to do the same concerning fictional *truth value gluts* as well. Extending both **L1** and **L2** to allow for inconsistent and incomplete fictional scenarios is one of the main issues we intend to address in future work.

Now, one might quite naturally ask whether a similar extension would also be required concerning the factual truth values, that is, whether there are (or may be) factual sentences that are neither factually true nor factually false or both. We shall not take a stance on this issue here, but merely point out that if the answer to this question were in the negative, that is, if the real world, unlike some of the "worlds" depicted in fiction, were assumed to be both consistent

[21] This situation also indicates a further possible implementation of the framework presented here, namely, to adopt different semantic clauses for different predicate letters depending on which natural language properties and relations they are supposed to express. On this proposal, for instance, At_1t_2 would be factually true if $\langle \overline{s}(t_1), \overline{s}(t_2)\rangle \in A^{\mathfrak{A}}$ and $\overline{s}(t_1) \in A$; and it would be fictionally true $\langle \overline{s}(t_1), \overline{s}(t_2)\rangle \in A^{\mathfrak{A}}$ and $\overline{s}(t_1) \in A_{\emptyset}$.

and complete, then the operator ! would have a rather interesting behavior. For it would allow us to restore some inferences that are usually invalid in contexts in which there are truth value gaps and gluts. Specifically, if truth value gaps were not allowed concerning the factual truth values, then we should expect the following inference schema to be valid:

$$\frac{!\alpha}{\alpha \vee \neg\alpha}$$

And if factual truth value gluts were not allowed, we should expect:

$$\frac{!\alpha \qquad \alpha \qquad \neg\alpha}{\beta}$$

to be valid as well. The resulting logic(s) would then belong to the family of systems known as *logics of formal inconsistency and undeterminedness* in the sense of [17], where the operator ! would simultaneously be a *consistency* and a *determinedness connective*.[22]

A third point tacitly assumed here is that a given object cannot be both fictional and real, which translates into our logical framework as the requirement that $A \cap A_{\emptyset} = \emptyset$. We acknowledge this assumption is somewhat controversial. For it seems reasonable to assume, by reading Conan Doyle's novel, that London as depicted in his stories is the same as the real London. You could say that it's not the same London as it is today, but at least it's the same London as when the novels were published. Note, however, that the sentence:

The building at 221B Baker St. London in 1887 exists (18)

is factually false, since that address only came to exist in London in 1930. We take this to be a compelling reason for not identifying the real and the fictional Londons. Moreover, if both cities were the same, what would be the truth value of (18)?

One possible response is to assign a pair of distinct truth values to each sentence. That is, every formula α would get assigned a pair $\langle V_1, V_2 \rangle$ such V_1 and V_2 are either T or F and represent respectively α's factual and fictional truth values. On this proposal, (18) would then receive the value $\langle F, T \rangle$ on its intended interpretation.[23] This semantics could even be modified to accommodate truth value gaps and gluts in the way suggested above, yielding four new pairs of values: $\langle T, N \rangle$, $\langle F, N \rangle$, $\langle T, B \rangle$, $\langle F, B \rangle$, where N and B stand respectively for *neither true nor false* and *both true and false*.

All the above suggestions are certainly worth pursuing to extend the initial framework presented in this paper.

[22] In other words, ! would be a *classicality operator* in the sense of [19, 23].

[23] Another way of implementing the same idea is to multiply the number of truth values by four. We would then have the values TT, Tt, Tf, TF, ..., and (18) would receive the value Ft.

Acknowledgments. We would like to thank the anonymous referees for valuable suggestions regarding earlier versions of this paper. The first author was supported by UFFS (Brazil), grant PES-2020-0134.

References

[1] Avron, A., and A. Zamansky, "Non-deterministic semantics for logical systems", pages 227–304 in *Handbook of Philosophical Logic*, vol. 16, 2011.

[2] Bacon, A., "Quantificational logic and empty names", *Philosophers' Imprint*, 13 (2013).

[3] Badura, C., and F. Berto, "Truth in fiction, impossible worlds, and belief revision", *Australasian Journal of Philosophy*, 97, 1 (2019): 178–193. DOI: 10.1080/00048402.2018.1435698

[4] Bencivenga, E., "Free logics", pages 147–196 in D. M. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2 ed., vol. 5, Springer, 2002.

[5] da Costa, N. C. A., "On the theory of inconsistent formal systems", *Notre Dame Journal of Formal Logic*, 15, 4 (1974): 497–510. DOI: 10.1305/ndjfl/1093891487

[6] Evans, G., "Understanding demonstratives", in *Collected Papers*, Oxford University Press, 1985.

[7] Everett, A., "Against fictional realism", *The Journal of Philosophy*, 102, 12 (2005): 624–649. DOI: 10.5840/jphil2005102129

[8] Frege, G., "Sense and reference", *The Philosophical Review*, 57 (1948): 209–230.

[9] Frege, G., *Posthumous Writings*, Blackwell, 1979.

[10] Gottwald, S., "Many-valued logic", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Summer 2020 edition, Stanford University. https://plato.stanford.edu/entries/logic-manyvalued/

[11] Ivlev, J. V., "A semantics for modal calculi", *Bulletin of the section of logic*, 17 (1988): 114–121.

[12] Jaquette, D. (ed.), *Philosophy of Logic*, 5 ed., *Handbook of the Philosophy of Logic*, North Holland, 2007.

[13] Kearns, J., "Modal semantics without possible worlds", *The Journal of Symbolic Logic*, 46, 1 (1981): 77–86. DOI: 10.2307/2273259

[14] Kripke, S. A., "Quantified modality and essentialism", *Noûs*, 51, 2 (2017): 221–234. DOI: 10.1111/nous.12126

[15] Lehmann, S., "Strict Fregean free logic", *Journal of Philosophical Logic*, 23, 3 (1994): 307–336. DOI: 10.1007/BF01048484

[16] Lehmann, S., "More free logic", pages 197–259 in D. M. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2 ed., vol. 5, Springer, 2002.

[17] Marcos, J., "Nearly every normal modal logic is paranormal", *Logique et Analyse*, 48 (2005): 279–300.

[18] Nolt, J., "Free logics", pages 1023–1060 in [12].

[19] Omori, H., and T. Waragai, "Some observations on the systems LFI1 and LFI1*", pages 320–324 in F. Morvan, A. M. Tjoa, and R. Wagner (eds.), *Proceedings of DEXA2011*, 2011, IEEE computer society.

[20] Orlando, L., "Fictional names without fictional objects", *Revista Hispanoamericana de Filosofía*, 40 (2008): 111–127.

[21] Priest, G., "Sylvan's box: a short story and ten morals", *Notre Dame Journal of Formal Logic*, 38, 4 (1997): 573–82. DOI: 10.1305/ndjfl/1039540770

[22] Sainsbury, R. M., *Reference Without Referents*, Oxford University Press, USA, 2005.

[23] Sano, K., and H. Omori, "An expansion of first-order Dunn-Belnap logic", *Logic Journal of the IGPL*, 22, 3 (2013): 458–481. DOI: 10.1093/jigpal/jzt044

[24] Sawyer, S., "The importance of fictional properties", pages 208–229 in S. Brock and A. Everett (eds.), *Fictional Object*, Oxford University Press, 2015. DOI: 10.1093/acprof:oso/9780198735595.003.0009

[25] Suppes, P., *Axiomatic Set Theory*, 2 ed., Dover, 1972.

[26] van Dalen, D., *Logic and Structure*, 5 ed., Springer, 2013. DOI: 10.1007/978-1-4471-4558-5

[27] Woods, J., "Fictions and their logic", pages 1061–1126 in [12].

Newton Peron
Department of Philosophy
Federal University of Southern Frontier (UFFS)
Chapecó, Brazil
newton.peron@gmail.com

Henrique Antunes
Department of Philosophy
Federal University of Minas Gerais (UFMG)
Belo Horizonte, Brazil
antunes.henrique@outlook.com