THEORIA

THEORIA, 2014
doi:10.1111/theo.12052

Solving Multimodal Paradoxes

by

FEDERICO PAILOS

and

LUCAS ROSENBLATT

University of Buenos Aires – CONICET

Abstract: Recently, it has been observed that the usual type-theoretic restrictions are not enough to block certain paradoxes involving two or more predicates. In particular, when we have a self-referential language containing modal predicates, new paradoxes might appear even if there are type restrictions for the principles governing those predicates. In this article we consider two type-theoretic solutions to multimodal paradoxes. The first one adds types for each of the modal predicates. We argue that there are a number of problems with most versions of this approach. The second one, which we favour, represents modal notions by using the truth predicate together with the corresponding modal operator. This way of doing things is not only useful because it avoids multimodal paradoxes, but also because it preserves the expressive capacity of the language. As an example of the sort of theory we have in mind, we provide a type-theoretic axiomatization that combines truth with necessity and knowledge.

Keywords: multimodal paradoxes, semantic paradoxes, modalities, truth, type-theory

1. Introduction

IT IS WELL KNOWN that the Liar and other self-referential paradoxes can be avoided by using typed theories. It is less known that for self-referential languages containing two or more modal predicates, new paradoxes might appear. Cases of this sort might be found in Halbach (2006, 2008), Horsten and Leitgeb (2001), Paseau (2009) and Stern and Fischer (2013). Moreover, it turns out that the usual type-theoretic restrictions are not enough to block these multimodal paradoxes. Discussing a paradox generated by a sentence saying of itself that it is unknowable, Volker Halbach (2008, p. 119) claims that:

> [t]he new paradox . . . belongs to a larger family of paradoxes that arise from the interaction of two or more intensional notions treated as predicates. After all, the typing of necessity and the typing of other notions are known to interact with lethal effects also in other cases . . . Therefore the paradox may highlight only a general problem with resolving paradoxes by typing. I am happy to concede this: there may be an underlying general problem with resolving paradoxes by typing, if more than one notion is under consideration.

Our goal is to show that the type theorist has the appropriate tools to solve these paradoxes and, moreover, that paradoxes of this sort pose no problems over and above the usual Liar-like paradoxes.

The article is structured as follows. In section 2 we provide some background regarding the distinction between predicates and operators in a formal framework, and we sketch how the usual paradoxes can be avoided by the introduction of types for the corresponding predicates. In section 3 we show how multimodal self-referential sentences are capable of generating inconsistencies even if type restrictions are present. In section 4 we provide two solutions to the paradoxes. The first one is considered in section 4.1, while the second one is informally explained in section 4.2. In section 5 we provide a simple axiomatic system to illustrate the sort of theory we want to support and we prove that it has some nice properties. Finally, section 6 deals very briefly with some objections that this sort of approach usually faces.

2. Predicates, Operators and the Type-Theoretic Approach

Nowadays, when working in a formal framework, it is quite common to treat some notions as operators and others as predicates. For example, while truth is generally treated as a predicate, necessity, knowledge, belief and other modalities are usually taken to be operators. On the one hand, the operator approach has the advantage that possible-world semantics has been rigorously developed and has proved to be extremely useful for notions such as necessity. On the other hand, the predicate approach has the advantage that the expressive resources of the language are suitable to make generalizations involving those predicates.

The problem with the approach that uses predicates is that when necessity, knowledge and other notions are represented by predicates, we can use them to generate modal paradoxes which are similar to the Liar.[1] The first relevant paradox we will discuss is due to Richard Montague.[2] He noticed that in order to arrive at an inconsistency in a theory that allows for self-reference it is not necessary, as in the Liar Paradox, to have the full strength of the unrestricted T-schema. Something weaker already leads to trouble.

(*Montague's Paradox*) Let $\mathcal{T}$ be a theory that extends Robinson's Arithmetic $\mathcal{Q}$ and let P(x) be a monadic predicate of the language $\mathcal{L}_{\mathcal{T}}$ of $\mathcal{T}$. If P(x) satisfies (P-elim) and (P-nec), then $\mathcal{T}$ is inconsistent.

$$\text{(P-elim)} \vdash_{\mathcal{T}} \mathrm{P}\ulcorner\phi\urcorner \rightarrow \phi.$$

$$\text{(P-nec) If } \vdash_{\mathcal{T}} \phi, \text{ then } \vdash_{\mathcal{T}} \mathrm{P}\ulcorner\phi\urcorner.$$ [3]

1 Although it is a bit non-standard, it is possible to obtain self-referential sentences without introducing predicates. Stern and Fischer (2013) investigate multimodal paradoxes in a setting where self-reference is obtained by the introduction of fixed point constants, a technique developed in Smorynski (2004).

2 See Montague (1963) for his version of the paradox.

3 Robinson's Arithmetic contains numerals that work as names of codes of expressions. The symbol $\ulcorner\ \urcorner$ works as a name-forming device. The expression $\ulcorner\phi\urcorner$ denotes (the code) of the sentence ϕ under some

For a long time the philosophical community took this last result as conclusive evidence that every concept P – except, curiously, for truth – satisfying (P-elim) and (P-nec) must be treated as an operator and not as a predicate. The point applies, for instance, to the notions of necessity and knowledge (if some assumptions about knowledge are made, in particular, only an idealized notion of knowledge satisfies (P-nec)). In fact, Montague himself claims about the former that "if necessity is to be treated syntactically, that is, as a predicate of sentences . . . then virtually all of modal logic . . . is to be sacrificed".[4]

But is this so? One rather popular way to solve the Liar Paradox is by introducing types. There are different ways of doing this within a theory of truth. One possibility is to syntactically ban all sentences of the form $T\ulcorner\phi\urcorner$ where ϕ contains an occurrence of the predicate T(x). Another possibility is to restrict the proof-theoretic principles in which the predicate T(x) is involved. For example, if we have a base theory $\mathcal{T}$ that extends $\mathcal{Q}$ and the predicate T(x) is added to the language $\mathcal{L}_{\mathcal{T}}$, instead of accepting the full unrestricted T-schema we only accept those instances of the T-schema in which ϕ is a formula of the language $\mathcal{L}_{\mathcal{T}}$.

In the same way, the introduction of types for P, whatever P may be, is useful to block any version of Montague's Paradox.[5] Assume again that we are working in a theory $\mathcal{T}$ that extends $\mathcal{Q}$ and we demand that only the following principles are in $\mathcal{T}$:

(P-elim*) $\vdash_{\mathcal{T}} P\ulcorner\phi\urcorner \rightarrow \phi$ if P does not occur in ϕ.

(P-nec*) If $\vdash_{\mathcal{T}} \phi$, then $\vdash_{\mathcal{T}} P\ulcorner\phi\urcorner$ if P does not occur in ϕ.

It is quite easy to check that a contradiction is no longer derivable in $\mathcal{T}$ even though we can find a sentence ϕ such that $\vdash_{\mathcal{T}} \phi \leftrightarrow \neg P\ulcorner\phi\urcorner$.

3. Multimodal Paradoxes

Although the type-theoretic approach works quite nicely for the usual self-referential paradoxes, a serious problem arises when two or more predicates satisfy principles like the ones above. The problem is that, within the sort of languages we

appropriate coding for the expressions of the language. In order to get the inconsistency, just use the following fact given by the Diagonal Lemma: $\vdash_{\mathcal{T}} \phi \leftrightarrow \neg P\ulcorner\phi\urcorner$.

4 See Montague (1963), p. 294.

5 So it seems quite strange that Tarski's solution to the Liar and Montague's solution to his paradox turned out to be so different.

are working with, predicates create opaque contexts. Let P_1 and P_2 be two different predicates. Consider the sentence:

$$P_1\ulcorner 0 = 0\urcorner \wedge P_2\ulcorner 0 \neq 1\urcorner$$

It is perfectly accurate to say that this sentence contains an occurrence of the predicate P_1 and an occurrence of the predicate P_2. Now consider the sentence:

$$P_2(\underset{\cdot}{P_1}\ulcorner 0 = 0\urcorner)^{6}$$

It would be false to say that this sentence contains an occurrence of P_1, since it only contains a term that, according to some fixed coding, stands for (the code) a sentence that contains an occurrence of P_1. This fact can be used to prove the following:

(*Multimodal Montague Paradox*) Let $\mathcal{T}$ be a theory that extends $\mathcal{Q}$ and let $P_1(x)$ and $P_2(x)$ be monadic predicates of the language $\mathcal{L}_\mathcal{T}$ satisfying (P_i-elim*) and (P_i-nec*) for $i = 1, 2$. It follows that $\mathcal{T}$ is inconsistent.

Proof. By the Diagonal Lemma, which is provable in $\mathcal{T}$, we know that $\vdash_\mathcal{T} \phi \leftrightarrow \neg P_2(\underset{\cdot}{P_1}\ulcorner\phi\urcorner)$. Now assume that $P_2(\underset{\cdot}{P_1}\ulcorner\phi\urcorner)$ holds. By applications of (P_2-elim*) and (P_1-elim*), we can obtain ϕ, which is equivalent to $\neg P_2(\underset{\cdot}{P_1}\ulcorner\phi\urcorner)$. By *Reductio*, we can infer that $\vdash_\mathcal{T} \neg P_2(\underset{\cdot}{P_1}\ulcorner\phi\urcorner)$. But then we have $\vdash_\mathcal{T} \phi$ and applications of (P_1-nec*) and (P_2-nec*) give us $\vdash_\mathcal{T} P_2(\underset{\cdot}{P_1}\ulcorner\phi\urcorner)$. Hence, we have a contradiction.

This inconsistency might be viewed as emerging from the interaction of any two (or more) modal predicates satisfying (P-elim*) and (P-nec*). Notice that this is a serious problem for the type theorist: the inconsistency is reachable even if the usual type restrictions are in place.[7]

The version of the paradox just given might strike the reader as too abstract. To stress its importance, we will provide a specific example that might be viewed as a particular case of the Multimodal Montague Paradox.

(*The Unknowability Paradox*)[8] Let $\mathcal{T}$ be a theory that extends $\mathcal{Q}$ and let $\Box(x)$ and $K(x)$ be predicates of $\mathcal{L}_\mathcal{T}$. If $\Box(x)$ and $K(x)$ satisfy the four rules below, then $\mathcal{T}$ is inconsistent.

---

6 We stipulate that in an expression like $P_2(\underset{\cdot}{P_1}\ulcorner\psi\urcorner)$, $\underset{\cdot}{P_1}: \omega \rightarrow \omega$ represents a function that when applied to (the code of) the sentence ψ outputs (the code of) the sentence $P_1\ulcorner\psi\urcorner$. In any case, it would do just as well to use the notation $P_2\ulcorner P_1\ulcorner 0 = 0\urcorner\urcorner$.

7 Of course, the paradox also threatens untyped theories, but this is not a big deal. The point is that the type theorist expects the usual type restrictions to avoid every sort of inconsistency. What the Multimodal Montague Paradox shows is that this demand cannot be fulfilled.

8 A version of this paradox can be found in Paseau (2009).

$$(\Box\text{-elim*}) \vdash_{\mathcal{T}} \Box\ulcorner\phi\urcorner \rightarrow \phi \quad \text{if } \Box \text{ does not occur in } \phi.$$

$$(\Box\text{-nec*}) \text{ If } \vdash_{\mathcal{T}} \phi \text{ then} \vdash_{\mathcal{T}} \Box\ulcorner\phi\urcorner \quad \text{if } \Box \text{ does not occur in } \phi.$$

$$(\mathrm{K}\text{-elim*}) \vdash_{\mathcal{T}} \mathrm{K}\ulcorner\phi\urcorner \rightarrow \phi \quad \text{if K does not occur in } \phi.$$

$$(\mathrm{K}\text{-nec*}) \text{ If } \vdash_{\mathcal{T}} \phi \text{ then} \vdash_{\mathcal{T}} \mathrm{K}\ulcorner\phi\urcorner \quad \text{if K does not occur in } \phi.$$

Proof. By the Diagonal Lemma, we have $\vdash_{\mathcal{T}} \phi \leftrightarrow \Box(\underset{\cdot}{\neg}\underset{\cdot}{\mathrm{K}}\ulcorner\phi\urcorner)$. Now assume that $\mathrm{K}\ulcorner\phi\urcorner$ holds. By (K-elim*) we can infer ϕ, and hence $\Box(\underset{\cdot}{\neg}\underset{\cdot}{\mathrm{K}}\ulcorner\phi\urcorner)$. By ($\Box$-elim*), we obtain $\neg\mathrm{K}\ulcorner\phi\urcorner$. So by *Reductio* we infer that $\vdash_{\mathcal{T}} \neg\mathrm{K}\ulcorner\phi\urcorner$. Now we can apply ($\Box$-nec*) to get $\vdash_{\mathcal{T}} \Box(\underset{\cdot}{\neg}\underset{\cdot}{\mathrm{K}}\ulcorner\phi\urcorner)$ which by the Diagonal Lemma is equivalent to ϕ. But if $\vdash_{\mathcal{T}} \phi$, by (K-nec*) also $\vdash_{\mathcal{T}}\mathrm{K}\ulcorner\phi\urcorner$, which gives us a contradiction.

There have been other multimodal paradoxes floating around in the literature.[9] For our present purposes, however, it is enough to consider the ones presented above.

4. Two Solutions

4.1 Interacting Types

We are going to consider two solutions to the multimodal paradoxes. The first one – which is actually a family of solutions – consists, to put it roughly, in adding more types. The second one, which we favour, consists in representing modal notions by using the truth predicate together with the corresponding modal operator. This solution will be presented in the next subsection.

Of course, it would be too extreme to ban all interactions between the relevant predicates. It would also be ineffective to use principles that are restricted to formulas that do not contain the relevant predicate *and that do not contain a term referring to a formula containing the relevant predicate*. The reason is that we can come up with Liar-like sentences that do not contain the relevant predicate nor a term referring to a formula containing the predicate, but that contain a term referring to a formula which in turn contains another term referring to another formula which does contain the relevant predicate. One might also try to use principles that are restricted to formulas that do not contain any explicit *or implicit*

9 A similar paradox, involving truth and necessity, can be found in Halbach (2006). The reader can also see Horsten and Leitgeb (2001) and Halbach (2008). These last two papers deal with multimodal paradoxes that, although quite interesting, are less problematic for the type theorist because they involve mixed principles, where, roughly, a mixed principle is one that, if expressed solely by means of operators, contains at least two different (kinds of) modal operators (for instance, one such principle would be the Knowability principle, according to which every truth can be known). But if that is the case, then the type theorist might argue that it is natural to demand mixed type restrictions.

occurrences of the relevant predicate. Unfortunately, as pointed out in by Halbach (2006), working out the details of this strategy is rather messy.

However, it is interesting to notice that the Multimodal Montague Paradox considered above can be blocked if cross-hierarchy restrictions are introduced. One fairly elegant way of doing this is by sequentially introducing infinitely many predicates $P_1, P_2, P_3, \ldots$ to the language of a base theory $\mathcal{T}$, where $\mathcal{T}$ is any sound arithmetical theory. Each of these predicates is supposed to represent a modality. The idea is that $\mathcal{L}_{\mathcal{T}} \cup \{P_1\}$ is $\mathcal{L}_{\mathcal{T}_{P_1}}$, $\mathcal{L}_{\mathcal{T}} \cup \{P_1, P_2\}$ is $\mathcal{L}_{\mathcal{T}_{P_1P_2}}$, and so on. For each new language $\mathcal{L}_{\mathcal{T}_{P_1\ldots P_n}}$ we have new restricted principles for P. The resulting language is $\mathcal{L}_{\mathcal{T}_{P_\omega}}$, that is, $\mathcal{L}_{\mathcal{T}} \cup \{P_1, P_1, P_3 \ldots\}$.[10] The resulting theory $\mathcal{T}_{P_\omega}$ has, for each $n > 0$, restricted versions of P-elim and P-nec. More formally, for each $n > 0$, the following holds:

$$(P_n\text{-elim*}) \vdash_{\mathcal{T}_{P_\omega}} \mathrm{P}_n\ulcorner\phi\urcorner \rightarrow \phi \text{ for each } \phi \in \mathcal{L}_{\mathcal{T}_{P_1\ldots P_{n-1}}}$$

$$(P_n\text{-nec*}) \text{ If } \vdash_{\mathcal{T}_{P_\omega}} \phi, \text{ then } \vdash_{\mathcal{T}_{P_\omega}} \mathrm{P}_n\ulcorner\phi\urcorner \text{ for each } \phi \in \mathcal{L}_{\mathcal{T}_{P_1\ldots P_{n-1}}}$$

If the typed predicates are introduced in this way, it is no longer possible to obtain a contradiction.[11]

Technically this solution works perfectly. Both the one-predicate paradoxes and the multimodal paradoxes are avoided. Conceptually, however, it is not easy to adequately motivate the order in which the different predicates are introduced. To see why, consider truth, necessity and knowledge. To make sure that the combination of these notions does not generate paradoxes, we should sequentially introduce infinitely many truth, necessity and knowledge predicates. But it is not clear how the different predicates are to be introduced. As a matter of fact, there are infinitely many ways of linearly ordering the predicates, but there is no good reason to prefer one order over another. They all seem unmotivated and artificial. Moreover, the artificiality grows as the number of new predicates in the language increases.[12]

There might be a reasonable way of *partially* ordering the predicates which does not involve the sort of artificiality we have discussed. For example, consider only

---

10 Obviously the hierarchy of predicates could continue at transfinite levels. And in fact, as one anonymous reviewer suggests, stopping at ω will render certain modal statements inexpressible in the language, such as the statement saying that all sentences of the form $\Box^{n}\ulcorner 0 = 0\urcorner$ are necessary (where $\Box^{n}$ stands for n occurrences of $\Box$). However, to keep things simple, we will only go as far as ω.

11 The reason is that now the predicates are linearly ordered. For example, the proof presented before for the Multimodal Montague sentence breaks down when we try to apply (P_1-elim*) to the sentence ϕ (and also in the step from ϕ to $P_1\ulcorner\phi\urcorner$ by the rule (P_1-nec*)).

12 As expected, we do not have a conclusive argument against all possible ways in which necessity, knowledge, truth and other notions could be linearly ordered. But in any case, it seems that the prospects for this sort of typing do not look good.

the necessity and the knowledge predicates. It will work to introduce two hierarchies for $\Box_n$ and K_n, with axioms and rules ($\Box_n$-elim*), (K_n-elim*), ($\Box_n$-nec*) and (K_n-nec*) such that for instance ($\Box_n$-nec*) says that $\vdash \Box_n\phi$ if $\vdash\phi$, and ϕ contains no instances of $\Box_k$, K_k, for $k \geq n$, and so on for the other rules. (This theory would be a fragment of one of the linear order theories we have described above, so it is consistent.) This approach will be further discussed in the last section.

4.2 Emulating Predicates

The approach that we are about to put forward does not have this sort of difficulties. It is based on the idea that we can emulate different predicates by using the truth predicate in combination with an operator. A similar approach has already been endorsed by some philosophers.[13] Our point is to stress that the approach provides the type theorist with the tools she needs to address multimodal paradoxes.[14]

Instead of representing a certain notion by a predicate P(x), we represent it by OT(x), where O is the operator that corresponds to the predicate P(x) and T(x) is the truth predicate. So in our proposal truth is treated as a predicate but all the other notions are treated as operators (as it is usually done).[15] The main advantage of this approach is that even if some notions are treated as operators, no expressive power is lost as long as the expressions of the form OT(x) are properly understood.[16] Notice however that we do not claim that OT(x) expresses the same concept as (or has the same meaning as) O. It is enough for our purposes that OT(x) behaves sufficiently like the modality presumably expressed by the operator O and, of course, that it avoids paradoxes.

Returning to necessity and knowledge, if O is $\Box$ we can think of the predicate $\Box$T(x) as partially capturing the concept of necessity and if O is K we can think of the predicate KT(x) as partially capturing the concept of knowledge. One might think it is unnatural to replace occurrences of P(x), whatever that is, by occurrences of the complex expression OT(x). But natural languages allow for paraphrases of

13 As a matter of fact the idea was sketched in Kripke (1975). See Halbach and Welch (2009) and Horsten (2009) for more recent applications of this idea.

14 The contents of this section and the next partially overlap with the work of one of the authors in Rosenblatt (2014). However, the emphasis there is on the Knowability Argument, whereas here we are only considering self-referential multimodal paradoxes.

15 We are not completely sure whether the *all* in the previous sentence should be understood unrestrictedly. Perhaps some notions (for instance, (non-idealized) belief and logical validity) can be treated as predicates without generating an inconsistency. It all depends, of course, on the principles each notion obeys.

16 The approach put forward here is similar in spirit to Halbach and Welch (2009). However, they are not immediately concerned with multimodal paradoxes but with the possibility of giving a translation of predicates in terms of operators (together with the truth predicate). Also, they offer a semantic account in which the meaning of the connectives is given by Strong Kleene logic and the extension (and the antiextension) of the truth predicate is given by the Kripke minimal fixed point.

this kind. For instance, if P(x) is the necessity predicate or the knowledge predicate, it seems plausible to paraphrase sentences like *2 + 2 = 4 is necessary* and *2 + 2 = 4 is known* by *2 + 2 = 4 is necessarily true* and *2 + 2 = 4 is known to be true*, respectively.[17]

5. A Theory of Truth, Necessity and Knowledge

To illustrate how our solution to the multimodal paradoxes works we will present a formal theory in which truth, necessity and knowledge can be represented. Naturally, our solution is meant to be general in two different ways. First, it is in principle capable of dealing with other modal notions as well. The main point of our approach is that certain notions (certainly knowledge and necessity, but maybe other notions too) need not be added as predicates if the truth predicate is already present. Second, we remain neutral as to whether the theory of truth we will use in this section is the right one. Our general approach is meant to be compatible with other theories of truth. If truth is the only notion treated as a predicate, whatever the solution to the truth paradoxes is, it will be useful to block the multimodal paradoxes, as long as the other notions are treated as operators. However, for the purpose of illustrating our point, the theory we are about to offer is more than enough.

Let $\mathcal{T}$ be a theory expressed in a first-order language $\mathcal{L}$ extending Peano Arithmetic. We stipulate that in addition to having the usual non-logical symbols of arithmetic the language $\mathcal{L}$ has finitely many *new* individual constants, functions, symbols and predicates.[18] As before, we assume that we have some appropriate form of coding for the expressions of the language and we take $\ulcorner e \urcorner$ to be the name of (the code of) the expression e. Let $\mathcal{L}_{\Box K}$ be the language that results from adding the necessity operator $\Box$ and the knowledge operator K to $\mathcal{L}$ (of course, we assume that $\mathcal{L}_{\Box K}$ also has names for expressions containing K and $\Box$). Let $\mathcal{T}_{\Box K}$ be the theory that results from adding the following axioms and rules to $\mathcal{T}$:

$$(\Box\text{-elim})\ \Box\phi \rightarrow \phi$$

$$(\Box\text{-nec})\ \text{If } \vdash_{\mathcal{T}_{\Box K}} \phi, \text{ then } \vdash_{\mathcal{T}_{\Box K}} \Box\phi$$

$$(\Box\text{-dist})\ \Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi)$$

17 It might be argued that the latter paraphrase is harder to justify, since an agent might know an arithmetical truth without possesing the concept of truth. We will consider this problem in section 6.3.

18 Since the whole point of the construction is to add modal operators to the language, we need to add resources to construct contingent sentences. Otherwise, the modal operators will collapse with truth.

$$(\Box\text{-trans})\ \Box\phi \to \Box\Box\phi$$

$$(\text{K-elim})\ \mathrm{K}\phi \to \phi$$

$$(\text{K-nec})\ \text{If } \vdash_{\mathcal{T}_{\Box K}} \phi, \text{ then } \vdash_{\mathcal{T}_{\Box K}} \mathrm{K}\phi$$

$$(\text{K-dist})\ \mathrm{K}(\phi \to \psi) \to (\mathrm{K}\phi \to \mathrm{K}\psi)$$ [19]

In other words, the theory $\mathcal{T}_{\Box K}$ has an S4-notion of necessity and a T-notion of knowledge (there is nothing special about this particular choice of axioms, they just keep the construction of models for the theory simple). As usual, the theory can be interpreted by a Kripke frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R}_{\Box}, \mathcal{R}_{K} \rangle$ such that $\mathcal{W}$ is a set of worlds, $\mathcal{R}_{\Box}$ is a reflexive and transitive accessibility relation on $\mathcal{W}$ and $\mathcal{R}_{K}$ is a reflexive accessibility relation on $\mathcal{W}$. We assume that the domain $\mathcal{D}$ of every world contains all the natural numbers and that no non-standard numbers are in it. In this way we make sure that whenever we use an appropriate interpretation based on this frame, the arithmetic part of the theory will be satisfied. Also, to avoid unnecessary complications and, more specifically, to simplify the interpretations, we are going to make three stipulations. First, we assume that the domain $\mathcal{D}$ is the same in every world.[20] Second, each individual constant denotes the same object in every world. So the only changes between worlds have to do with the interpretation of (some of) the function symbols and (some of) the predicates.[21] Third, the language has enough names for the objects in $\mathcal{D}$. So we can develop the semantics for the quantifiers substitutionally.[22]

It is not hard to show that we can use this frame $\mathcal{F}$ to build a model $\mathcal{M}$ based on $\mathcal{F}$ for $\mathcal{T}_{\Box K}$. The models for $\mathcal{T}_{\Box K}$ are of the form $\langle \mathcal{W}, \mathcal{R}_{\Box}, \mathcal{R}_{K}, \mathcal{I}^{w} \rangle$ where $\mathcal{W}$, $\mathcal{R}_{\Box}$ and $\mathcal{R}_{K}$ are as before and $\mathcal{I}^{w}$ is a function that assigns in each world $w \in \mathcal{W}$ an object $c^{\mathcal{M}}$ in $\mathcal{D}$ to each individual constant *c*, a function $f^{\mathcal{M}}$ from $\mathcal{D}^{k}$ to $\mathcal{D}$ to each k-ary function symbol *f* and a relation $r^{\mathcal{M}}$ in $\mathcal{D}^{k}$ to each k-ary relation

19 In these schemata, ϕ and ψ are meant to be closed formulae.
20 A consequence of this is that the Barcan formulas for K and $\Box$ are true in every interpretation. But of course it is perfectly possible to set things differently so that they are no longer universally valid.
21 Two observations. First, this simplification makes the following formulae true in every structure:

$$(\Box\text{-ineq})\ \mathrm{t}_1 \neq \mathrm{t}_2 \to \Box(\mathrm{t}_1 \neq \mathrm{t}_2) \text{ for rigid terms } \mathrm{t}_1 \text{ and } \mathrm{t}_2$$

$$(\text{K-ineq})\ \mathrm{t}_1 \neq \mathrm{t}_2 \to \mathrm{K}(\mathrm{t}_1 \neq \mathrm{t}_2) \text{ for rigid terms } \mathrm{t}_1 \text{ and } \mathrm{t}_2$$

(We say that a term t is *rigid* if and only if the denotation of t is the same at every world). Again, things could have been set otherwise. Second, some function symbols (such as successor, addition and multiplication) will of course denote the same function in every world.
22 Once again, nothing crucial depends on these simplifications. They just make the construction of models for the theory much easier.

symbol r. We assume that $\mathcal{I}^w$ interprets the arithmetical vocabulary in the standard way. An assignment of truth values to the formulas of $\mathcal{L}_{\Box K}$ relative to a model $\mathcal{M}$ and a world ω can be defined straightforwardly. For example, we say that $(\mathcal{M}, w) \vDash r(\mathrm{t}_1, \ldots, \mathrm{t}_k)$ if and only if $\langle \mathcal{I}^w(\mathrm{t}_1), \ldots, \mathcal{I}^w(\mathrm{t}_k)\rangle \in \mathcal{I}^w(r)$, and that $(\mathcal{M}, w) \vDash \Box\phi$ if and only if $(\forall v \in \mathcal{W})(w\mathcal{R}_{\Box}v \rightarrow (\mathcal{M}, v) \vDash \phi)$.

It can be shown that there are assignments of truth values that make all the axioms and rules for K and □ true. As a matter of fact, any model $\mathcal{M}$ based on the frame $\mathcal{F}$ with the features mentioned above will make all these principles true and all the rules valid.[23] So far, so good.

The difficult part is to add a truth predicate (or a sequence $\mathrm{T}_1(\mathrm{x})$, $\mathrm{T}_2(\mathrm{x})$, $\mathrm{T}_3(\mathrm{x})$, . . . of truth predicates) to $\mathcal{L}_{\Box K}$ and principles for truth to $\mathcal{T}_{\Box K}$. Let $\mathcal{L}_{\Box KT_1(x)}$ be the language that results from adding the truth predicate $\mathrm{T}_1(\mathrm{x})$ to $\mathcal{L}_{\Box K}$, $\mathcal{L}_{\Box KT_2(x)}$ the language that results from adding the truth predicates $\mathrm{T}_1(\mathrm{x})$ and $\mathrm{T}_2(\mathrm{x})$ to $\mathcal{L}_{\Box K}$, and so on. Finally, let $\mathcal{L}_{\Box KT_\omega(x)}$ be the language that results from adding the sequence of truth predicates $\mathrm{T}_1(\mathrm{x})$, $\mathrm{T}_2(\mathrm{x})$, $\mathrm{T}_3(\mathrm{x})$, . . . to $\mathcal{L}_{\Box K}$. $\mathcal{L}_{\Box KT_\omega(x)}$ contains a predicate $\mathrm{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(\mathrm{y})$ that is true of exactly the (codes of the) sentences of $\mathcal{L}_{\Box KT_n(x)}$ and a function symbol $val^{\omega}(\mathrm{x})$ that, when applied to a closed term of $\mathcal{L}_{\Box KT_n(x)}$, gives the value of that term at world ω.[24] Let $\mathcal{T}_{\Box KT_\omega(x)}$ be the theory that results from adding the following axioms to $\mathcal{T}_{\Box K}$ for each $n \in \omega$ and for each predicate r of $\mathcal{L}_{\Box KT_\omega(x)}$:

$$(\mathrm{At}_{n+1})\ \forall \mathrm{t}_1, \ldots, \forall \mathrm{t}_k(\mathrm{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(\underset{\cdot}{r}(\mathrm{t}_1, \ldots, \mathrm{t}_k)) \rightarrow (\mathrm{T}_{n+1}(\underset{\cdot}{r}(\mathrm{t}_1, \ldots, \mathrm{t}_k)) \leftrightarrow r(val^w(\mathrm{t}_1), \ldots, val^w(\mathrm{t}_k))))^{25}$$

$$(\mathrm{Comp}\neg_{n+1})\ \forall \mathrm{y}(\mathrm{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(\mathrm{y}) \rightarrow (\mathrm{T}_{n+1}(\underset{\cdot}{\neg}\mathrm{y}) \leftrightarrow \neg\mathrm{T}_{n+1}(\mathrm{y})))$$

$$(\mathrm{Comp}\wedge_{n+1})\ \forall \mathrm{y}\forall \mathrm{z}(\mathrm{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(\mathrm{y} \underset{\cdot}{\wedge} \mathrm{z}) \rightarrow (\mathrm{T}_{n+1}((\mathrm{y} \underset{\cdot}{\wedge} \mathrm{z})) \leftrightarrow \mathrm{T}_{n+1}(\mathrm{y}) \wedge \mathrm{T}_{n+1}(\mathrm{z})))$$

$$(\mathrm{Comp}\exists_{n+1})\ \forall \mathrm{v}\forall \mathrm{y}(\mathrm{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(\underset{\cdot}{\exists}\mathrm{vy}) \rightarrow (\mathrm{T}_{n+1}(\underset{\cdot}{\exists}\mathrm{vy}) \leftrightarrow \exists \mathrm{t}(\mathrm{T}_{n+1}(\mathrm{y}(\mathrm{t}/\mathrm{v})))))^{26}$$

---

23 Actually things are not so simple. Since we have non-rigid terms in the language, the principles

$$\phi(\mathrm{t}) \rightarrow \exists \mathrm{x}\phi \mathrm{x}$$
$$\mathrm{s} = \mathrm{t} \rightarrow (\phi(\mathrm{s}) \leftrightarrow \phi(\mathrm{t}))$$

must be restricted to rigid terms s and t. Otherwise it would be easy to come up with counterexamples. For the first, let $\phi(\mathrm{x})$ be "it is necessary that x is tall" and let t be "the president of Argentina"; and for the second let $\phi(\mathrm{x})$ be "it is known that the morning star is x", let s be "the morning star" and let t be "the evening star". For more details, see Fagin et al. (2003), ch. 3.

24 We will assume that the function $\mathrm{val}^{\omega}(\mathrm{x})$ is representable in $\mathcal{T}_{\Box KT_\omega(x)}$.

25 In this axiom we use the dot notation in a somewhat different way. In $\mathrm{T}_{n+1}(\underset{\cdot}{r}(\mathrm{t}_1, \ldots, \mathrm{t}_k))$, $\underset{\cdot}{r}: \omega^k \rightarrow \omega$ represents a function that when applied to (the tuple of codes of) *the terms* $\mathrm{t}_1, \ldots, \mathrm{t}_k$ outputs (the code of) the sentence $r(\mathrm{t}_1, \ldots, \mathrm{t}_k)$.

26 Since we have stipulated that we have a term for every object, there is no need to use a satisfaction predicate.

$$(\Box \mathrm{T}_{n+1})\ \forall \mathrm{y}(\mathrm{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(\mathrm{y}) \rightarrow (\mathrm{T}_{n+1}(\underset{\cdot}{\Box}\mathrm{y}) \leftrightarrow (\Box \mathrm{T}_{n+1}(\mathrm{y})))$$

$$(\mathrm{KT}_{n+1})\ \forall \mathrm{y}(\mathrm{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(\mathrm{y}) \rightarrow (\mathrm{T}_{n+1}(\underset{\cdot}{\mathrm{K}}\mathrm{y}) \leftrightarrow \mathrm{KT}_{n+1}(\mathrm{y})))$$

$$(\mathrm{TT}_{n+1})\ \forall \mathrm{t}(\mathrm{Sent}_{\mathcal{L}_{\Box KT_{j-1}(x)}}(\mathrm{val}^{w}(\mathrm{t})) \rightarrow (\mathrm{T}_{n+1}(\underset{\cdot}{\mathrm{T}_j}(\mathrm{t})) \leftrightarrow (\mathrm{T}_{n+1}(\mathrm{val}^{w}(\mathrm{t})))) \text{ for each } j \leq n^{27}.$$

Clearly, the truth axioms are restricted to the T_{n+1}-free part of the language (but they can contain occurrences of $\Box$, K and $\mathrm{T}_j(\mathrm{x})$ for $j < n + 1$). Of course, we need seven such axioms for each truth predicate in the language, so the theory $\mathcal{T}_{\Box KT_\omega(x)}$ contains infinitely many (At_{n+1}) axioms, infinitely many ($\mathrm{Comp}\neg_{n+1}$) axioms, and so on. Also, once the truth predicates are part of the language, we can strengthen the schemata presented above for $\Box$ and K by replacing them with their universally quantified versions. For example, instead of ($\Box$-elim) the theory will contain for each n, an axiom of the form

$$\mathcal{F} = \langle \mathcal{W}, \mathcal{R}_\Box, \mathcal{R}_K \rangle$$

Let $\mathcal{F} = \langle \mathcal{W}, \mathcal{R}_\Box, \mathcal{R}_K \rangle$ be the theory that results from adding each of these axioms to Peano Arithmetic. Observe that this theory does not yet contain axioms that govern the interaction of truth with other logical expressions, although the language of the theory does contain truth predicates.[28]

In order to establish the consistency of the theory $\mathcal{T}_{\Box KT_\omega(x)}$ and to show that its modal operators can be interpreted in a plausible way, we now construct a model for it. A model $\mathcal{M}$ for the theory $\mathcal{T}_{\Box KT_\omega(x)}$ will be a structure of the form $\langle \mathcal{W}, \mathcal{R}_\Box, \mathcal{R}_K, \mathcal{I}^w, \mathcal{EXT} \rangle$, where $\mathcal{W}, \mathcal{R}_\Box, \mathcal{R}_K, \mathcal{I}^w$ are as before, and $\mathcal{EXT}: \mathcal{W} \times \{\mathrm{T}_n(\mathrm{x}): n \in \omega\} \rightarrow \mathcal{P}(\mathrm{Sent}_{\mathcal{L}_{\Box KT_n(x)}})$ is a function that assigns an extension to each truth predicate at each world $w \in \mathcal{W}$ and that respects the following condition:

For every $w \in \mathcal{W}$ and every ϕ in $\mathcal{L}_{\Box KT_n(x)}$, $(\mathcal{M}, w, \mathcal{EXT}) \vDash \mathrm{T}_{n+1}\ulcorner \phi \urcorner$ if and only if $(\mathcal{M}, w, \mathcal{EXT}) \vDash \phi$.

Here the notation "$(\mathcal{M}, w, \mathcal{EXT}) \vDash \phi$" is used to represent the fact that the formula ϕ is true relative to the model $\mathcal{M}$, the world ω and the function $\mathcal{EXT}$. It is unproblematic to assume that the theory $\mathcal{T}^{+}_{\Box K}$ has models of the form

27 It would be redundant to add axioms of the form

$$(\mathrm{TT}'_{n+1})\ \forall \mathrm{t}(\mathrm{Sent}_{\mathcal{L}_{\Box KT_{j-1}(x)}}(\mathrm{val}^{w}(\mathrm{t})) \rightarrow (\mathrm{T}_{n+1}(\underset{\cdot}{\mathrm{T}_j}(\mathrm{t})) \leftrightarrow (\mathrm{T}_j(\mathrm{val}^{w}(\mathrm{t}))))$$

since they are cases of (At_{n+1}).

28 Since we are not interested here in the issue of the conservativeness of our theory over Peano Arithmetic, we intentionally leave open whether the truth predicates can occur in formulas in the induction schema.

$\langle \mathcal{W}, \mathcal{R}_{\Box}, \mathcal{R}_K, \mathcal{I}^w \rangle$. For instance, it is not difficult to see that the reflexivity of $\mathcal{R}_{\Box}$ guarantees the truth of

$$\forall y(\text{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(y) \to (\Box T_{n+1}(y) \to T_{n+1}(y))).$$

It remains to show that the new axioms for truth are also true in this structure (for each n). We just present the proofs for the novel axioms ($\Box T_{n+1}$), (KT_{n+1}) and (TT_{n+1}).

Proof. For ($\Box T_{n+1}$) assume that for some model $\mathcal{M}$ (with the described features), world w and function $\mathcal{EXT}$, $(\mathcal{M}, w, \mathcal{EXT}) \vDash T_{n+1}(\dot{\Box}\ulcorner\phi\urcorner)$ for some formula $\phi \in \text{Sent}_{\mathcal{L}_{\Box KT_n(x)}}$. By the definition of $\mathcal{EXT}$, that holds if and only if $(\mathcal{M}, w, \mathcal{EXT}) \vDash \Box\phi$. Applying the semantic clause for $\Box$, we know that the previous fact holds if and only if $(\mathcal{M}, v, \mathcal{EXT}) \vDash \phi$ for each world v such that $wR_{\Box}v$. This is the case if and only if $(\mathcal{M}, v, \mathcal{EXT}) \vDash T_{n+1}\ulcorner\phi\urcorner$ for each world v such that $wR_{\Box}v$. In turn, by the semantic clause for $\Box$, this holds if and only if $(\mathcal{M}, w, \mathcal{EXT}) \vDash \Box T_{n+1}\ulcorner\phi\urcorner$.
For (KT_{n+1}) just replace in the previous paragraph $\Box$ everywhere by K.
For (TT_{n+1}), assume that for some model $\mathcal{M}$ (with the described features), world w and function $\mathcal{EXT}$, $(\mathcal{M}, w, \mathcal{EXT}) \vDash T_{n+1}(\dot{T}_j(t))$ for some term t such that $val^w(t) = \phi$ for some formula $\phi \in \text{Sent}_{\mathcal{L}_{\Box KT_{j-1}(x)}}$. By the definition of $\mathcal{EXT}$ this holds if and only if $(\mathcal{M}, w, \mathcal{EXT}) \vDash T_j\ulcorner\phi\urcorner$. And this is the case if and only if $(\mathcal{M}, w, \mathcal{EXT}) \vDash \phi$, which holds if and only if $(\mathcal{M}, w, \mathcal{EXT}) \vDash T_{n+1}\ulcorner\phi\urcorner$.

This is enough to establish:

Theorem 5.1. *$\mathcal{T}_{\Box KT_\omega(x)}$ is consistent.*

Observe that $\mathcal{T}_{\Box KT_\omega(x)}$ can be shown to be consistent without appealing to a model-theoretic construction like the one we just provided. The modal operators $\Box$ and K can be interpreted as double negations. So the modal axioms introduced above turn into logical truths. For instance, ($\Box$-elim) becomes $\neg\neg\phi \to \phi$. As for the axioms ($\Box T_{n+1}$) and (KT_{n+1}), they become

$$\forall y(\text{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(y) \to (T_{n+1}(\dot{\neg}\dot{\neg}y) \leftrightarrow \neg\neg T_{n+1}(y)))$$

and it is straightforward that this sentence follows from ($\text{Comp}\neg_{n+1}$). And in fact, if these operators are interpreted as double negations (or as the empty operator), then the usual model for the Tarski hierarchy up to ω is enough to prove consistency (and ω-consistency as well).

However, if so, as two anonymous reviewers pointed out to us, it could be argued that $\Box$ and K do not really add new things. But naturally, the model construction provided above is meant to show more than just mere consistency. In particular, it shows that we can interpret the modal operators in a plausible way. For instance, since there are contingent sentences available in the language, the construction shows that it is possible to express that certain sentences are true but not necessarily true, or true but not known to be true, among other things. In other words, the theory has models that enjoy all the desirable properties of possible world semantics.[29]

It can be shown that $\mathcal{T}_{\Box KT_\omega(x)}$ has other attractive features as well. For example, we can recover the T-schema for well behaved sentences:

Theorem 5.2. *If a sentence ϕ has no occurrences of $T_j(x)$ for $j > n$, then* $\vdash_{\mathcal{T}_{\Box KT_\omega(x)}} T_{n+1}\ulcorner\phi\urcorner \leftrightarrow \phi$ *for each $n \in \omega$.*

Proof. This can be proved by a simple induction on the complexity of ϕ. The base step is straightforward and it includes the case where ϕ is of the form $T_j\ulcorner\psi\urcorner$ for $j \leq n$. For the inductive step we have five cases. The formula ϕ can be of each of the following forms: $\neg\psi$, $\psi_1 \wedge \psi_2$, $\exists v\psi$, $\Box\psi$ and $K\psi$. We skip the usual cases and go straight to the modalities. If ϕ is of the form $\Box\psi$, by the inductive hypothesis we know that $\vdash_{\mathcal{T}_{\Box KT_\omega(x)}} T_{n+1}\ulcorner\psi\urcorner \leftrightarrow \psi$. From this it follows by ($\Box$-nec) and ($\Box$-dist) that $\vdash_{\mathcal{T}_{\Box KT_\omega(x)}} \Box T_{n+1}\ulcorner\psi\urcorner \leftrightarrow \Box\psi$. Also by ($\Box T_{n+1}$) we know that $\vdash_{\mathcal{T}_{\Box KT_\omega(x)}} T_{n+1}(\underset{\cdot}{\Box}\ulcorner\psi\urcorner) \leftrightarrow \Box T_{n+1}\ulcorner\psi\urcorner$. But then logical reasoning gives us $\vdash_{\mathcal{T}_{\Box KT_\omega(x)}} T_{n+1}(\underset{\cdot}{\Box}\ulcorner\psi\urcorner) \leftrightarrow \Box\psi$. (If ϕ is of the form $K\psi$, the proof is similar.)

It is clear that the T-schema does not hold unrestrictedly. For instance, $T_n\ulcorner\phi\urcorner \leftrightarrow \phi$ is not a theorem of $\mathcal{T}_{\Box KT_\omega(x)}$ if ϕ contains an occurrence of $T_n(x)$. This shows that multimodal paradoxes are no longer a problem. Although a sentence such as

$$\phi \leftrightarrow \neg\Box T_n(\underset{\cdot}{K}\,\underset{\cdot}{T}_n\ulcorner\phi\urcorner)$$

is provable in the theory using the Diagonal Lemma, no contradiction follows from it, since ϕ contains an occurrence of $T_n(x)$.

As a corollary of the preceding theorem we obtain the following nice interaction result between truth and necessity on the one hand, and truth and knowledge on the other:

Corollary 5.3. *If a sentence ϕ has no occurrences of $T_j(x)$ for $j > n$, then* $\vdash_{\mathcal{T}_{\Box KT_\omega(x)}} \Box T_{n+1}\ulcorner\phi\urcorner \leftrightarrow \Box\phi$, and $\vdash_{\mathcal{T}_{\Box KT_\omega(x)}} KT_{n+1}\ulcorner\phi\urcorner \leftrightarrow K\phi$.

This means that for non-problematic sentences, the necessity operator and the knowledge operator are equivalent to any predicate of the form $\Box T_{n+1}(x)$ and $KT_{n+1}(x)$, respectively.

29 Thanks to an anonymous reviewer for this point.

6. Four Objections

6.1 Is This a Type-Theoretic Solution to the Multimodal Paradoxes?

We have set up a contrast between type-theoretic solutions to paradoxes of self-reference and operator-theoretic solutions. Then we have sketched, in defence of the type-theoretic approach, a formalism that incorporates both operators and typed predicates. But it might be argued that the type theory is not doing any work in blocking the multimodal paradoxes. It is not as if we have an inconsistent theory of necessity and knowledge, which we make consistent by introducing type-distinctions. Instead we have a consistent theory of necessity and knowledge, in which we have already, by treating these notions as operators, blocked for instance the unknowability paradox discussed in section 3. Then we add typed truth-predicates to this theory, which we can usually do unproblematically. So how does our theory amount to a type-theoretic solution to multimodal paradoxes like the unknowability paradox?

It is not entirely true that by treating every modal notion as an operator, it is no longer possible to express multimodal sentences that might lead to paradoxes, especially if the language contains a truth-like predicate. Consider again the unknowability paradox. This is caused by a sentence saying of itself that it is unknowable. In the language of the theory we have sketched in the previous section, this sentence can be expressed by ϕ, where $\phi \leftrightarrow \Box T_n(\underset{\cdot}{\neg}\underset{\cdot}{K}\underset{\cdot}{T}_n\ulcorner\phi\urcorner)$ is a theorem of the theory. This sentence involves more than one modality (although we must grant that one of the modalities is only represented by a function symbol which is part of a singular term which names a sentence that contains the modality). If the truth predicate were not typed, a contradiction would be derivable. Moreover, it is clear that this sentence is not equivalent to any paradox-generating unimodal sentence, since the derivation of a contradiction from the former (but not from the latter) would require the relevant modal principles. So it is a bit misleading to say that the way in which the multimodal paradoxes are solved is by making them inexpressible in the language.[30] It might be said in response that this is not the kind of multimodal sentence that was considered previously, where the truth predicate played no role. But these complex expressions involving the truth predicate are precisely one way in which modalities can be represented in our language (see section 6.3).

30 The objection might rest in the way we have presented our theory in section 5. There we first introduce a theory for the modal operators K and $\Box$, and then give principles for truth on top of that theory. But this is just an elegant way of doing things. There is no real priority of the modal operators over truth. We could start by giving axioms for truth and then introducing the modal principles. The resulting theory would be equivalent.

6.2 Problems with Partially Ordering the Predicates

At the end of section 4.1 we suggested that there is a way of partially ordering the modal predicates which is not subject to the objections we raised against all ways of linearly ordering them. Then we went on to introduce the approach that emulates modal notions by using the truth predicate. Although we find no substantive advantage of the latter approach over the former, we believe that the former suffers from two drawbacks. The first one is related to the sort of justifications usually provided to type knowledge. The types are supposed to represent the way in which a particular proposition has come to be known by an agent. In Paseau's words (2008, p. 163): "the typing is one of epistemic access rather than (just) content". The following example is provided in Paseau (2009, p. 284): "I know2 that Giggs scored the winning penalty in the 2008 Champions League final if I infer it from the facts that you know1 that it was either Giggs or Ronaldo, and that Ronaldo missed his penalty". Also Linsky (2009, p. 172), subscribes to a similar line of reasoning: "We need not in general accept the principle $K_2p \rightarrow K_1p$. . . . Think of an . . . agent developing beliefs in order . . . There is no reason to believe that . . . what is known at a higher level must be known at lower levels. . . . What is known is frequently a function of other beliefs and knowledge". Now if there are other predicates like necessity (possibility) involved, the type theorist might want to make the necessity (possibility) types interact with the knowledge types, claiming that it is strange to have different and isolated hierarchies for knowledge and necessity (possibility). As a matter of fact, Paseau (2009, p. 282) argues that "it is natural to assume [that] there are such restrictions". Unfortunately, it is important to notice that nothing of what was said to justify the typing of the knowledge predicate seems to apply to possibility or necessity. Moreover, there is no non-ad hoc reason to think that the necessity types are somehow related to the knowledge types. The idea that the knowledge types represent the way in which knowledge is acquired by an agent cannot be transferred to the alethic modalities. It makes no sense to talk about "the way in which something has become necessary". So, in the case of necessity (possibility), the types, if there are any, should be content-based rather than access-based. Hence, the hierarchy of necessity (possibility) predicates and the hierarchy of knowledge predicates should presumably look different. This means that the approach sketched in section 4.1 seems unmotivated, given that it treats the necessity types exactly in the same way as it treats the knowledge types.

The second drawback of this approach follows from the way in which knowledge types are justified, and is related to how each approach deals with the Knowability Argument (sometimes also called the Fitch-Church Paradox).[31] According to this argument, we cannot reject the unreasonable claim that every truth is known if we

31 For an excellent discussion on the Knowability Argument see the papers in Salerno (2009).

accept the (not so unreasonable) claim that every truth is knowable. While our approach leaves the soundness of this argument untouched, the approach considered in section 4.1 automatically blocks it by the introduction of knowledge types. Recall that the argument assumes the sentence

$$\mathrm{K}(\phi \wedge \neg \mathrm{K}\phi).$$

In virtue of the type restrictions for knowledge, assuming that the knowledge type of ϕ is less than n, we would only have (for $m > n$) the following:

$$\mathrm{K}_m(\phi \wedge \neg K_n\phi).$$

From this we can only infer that

$$\mathrm{K}_m\phi \wedge K_m\neg K_n\phi.$$

But this is far from being a contradiction. A contradiction would be obtainable if the following principle holds for every m and n such that $m > n$:

$$\mathrm{K}_m\phi \rightarrow \mathrm{K}_n\phi.$$

But this is not acceptable according to the interpretation of the knowledge types provided above. An agent might know ϕ at type m without knowing it at a lower type n. Hence, there is no way to carry out the Knowability Argument in this approach. Of course, if the Knowability Argument is indeed a genuine paradox, this might be viewed as a virtue. However, we believe this is implausible for two reasons. First, many philosophers find the argument an interesting and convincing *reductio* of the claim that every truth is knowable (a claim usually associated with several kinds of anti-realist position). Second, although we fully agree with the claim that whenever we deal with two paradoxes of the same kind, they should receive the same form of answer, this seems not to be the case here. We see no reason to think that the Knowability Argument belongs to the same family of paradoxes as the self-referential (unimodal or multimodal) paradoxes. If we are right, then it seems undesirable (or strange at least) that considerations regarding self-referential paradoxes affect the debate over the plausibility of the claim that every truth is knowable.[32]

6.3 Knowing that ϕ and Knowing that ϕ Is True

It could be argued that the second fact in Theorem 5.3 is simply false because an agent can know ϕ without knowing that ϕ is true. This could happen for two

32 Similar considerations are put forward by one of the authors in Rosenblatt (2014).

different reasons. Either because the agent lacks the concept of truth or because the agent does not know what proposition is expressed by the sentence named by $\ulcorner\phi\urcorner$.[33] We do not make a lot of this. The framework we are providing is intended to represent an idealized notion of knowledge. This sort of idealization is absolutely standard in epistemic logic. The agent is assumed to be logically omniscient. Dropping this assumption means moving to the far less familiar territory of non-normal epistemic logics. We do not have objections to this move in principle, but here we are only interested in an idealized notion of knowledge. Now, since our theory is not just about knowledge but also about truth, and since truth requires the presence of names for the sentences in the language, we require two further idealizations. First, we assume that the agent (whose reasoning about knowledge we are theorizing about) has the concept of truth. Second, we assume that the agent knows what proposition is expressed by the sentence named by $\ulcorner\phi\urcorner$. We do not take these idealizations to be much stronger than the previous one. If we are theorizing about truth, it seems completely reasonable to assume that the agents have the corresponding concept. And if the language contains both an operator K (that applies to sentences) and a predicate KT(x) (that applies to sentence names), it also seems reasonable to demand that the first applies exactly to the sentences named by the terms to which KT applies.

This might strike the reader as an inadmissible idealization. In any case, the theory can be easily tweaked to handle the objection. Instead of having axioms of the form

$$(\mathrm{KT}_{n+1})\ \forall y(\mathrm{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(y) \rightarrow (\mathrm{T}_{n+1}(\underset{\cdot}{\mathrm{K}}y) \leftrightarrow \mathrm{KT}_{n+1}(y)))$$

we can have only the right to left direction:

$$(\mathrm{KT}_{n+1}{}')\ \forall y(\mathrm{Sent}_{\mathcal{L}_{\Box KT_n(x)}}(y) \rightarrow (\mathrm{KT}_{n+1}(y) \rightarrow \mathrm{T}_{n+1}(\underset{\cdot}{\mathrm{K}}y))).$$

This gives us a restricted version of Theorem 5.2:

Theorem 6.1. *If a sentence ϕ has no occurrences of $T_j(x)$ for $j > n$ and no occurrences of K, then $\vdash_{\mathcal{T}_{\Box KT_\omega(x)}} T_{n+1}\ulcorner\phi\urcorner \leftrightarrow \phi$ for each $n \in \omega$.*

Proof. The proof is similar to that of Theorem 5.2.

So Corollary 5.3 no longer holds. KT and K are not intersubstitutable in every context. However, we still have:

$$\vdash_{\mathcal{T}_{\Box KT_\omega(x)}} \mathrm{KT}_{n+1}\ulcorner\phi\urcorner \rightarrow \mathrm{K}\phi.$$

33 Although it is only the second problem that is usually discussed in the literature, we think the first problem is just as relevant.

We lose only the other direction, which was the direction the objection was aiming at.

6.4 The Asymmetry

A natural objection to our approach is that the symmetry between truth and the other notions is lost, since the first is treated as a predicate while the others as operators. This might be seen as a problem because Montague's paradox and the Liar paradox are reasonably said to belong to the same family of paradoxes, and it is usually acknowledged that whenever we face the same kind of paradox, we should provide the same kind of solution.[34]

However, notice that in this framework there is a sense in which all self-referential paradoxes do receive the same kind of solution. The restrictions we have imposed on the principles governing the truth predicate are useful not only to block the truth-theoretic paradoxes but also the unimodal and multimodal paradoxes. This is, we think, an advantage over the account considered in objection 4.2, where the truth-theoretic paradoxes require types for truth, the knowledge paradoxes require types for knowledge, and so on. So while it is true that in our approach there is an asymmetry in the way of representing truth and the other modal notions, there is no asymmetry regarding the sort of solution we provide to paradoxes belonging to the same family.[35]

Acknowledgements

The authors are very grateful to two anonymous reviewers of *Theoria* for their valuable comments and suggestions. We would also like to thank Lavinia Picollo for her extensive corrections to a previous version of this work. We are indebted to several members of the Buenos Aires Logic Group, especially Eduardo Barrio, Natalia Buacar, Eleonora Cresto, Damián Szmuc, Diego Tajer and Paula Teijeiro, for their collaboration. We would also particularly like to thank Catrin Campbell-Moore, Roy Cook, and Johannes Stern for their suggestions on how to improve the article. Both authors would like to stress that the article could not have been written without the financial aid of CONICET (Consejo National de Investigaciones Científicas y Tecnológicas).

34 Naturally, we do not have a rigorous account of what it means to belong to the same family, but we take it that any decent account would classify Montague's paradox and the Liar paradox as belonging to the same family.

35 Another possible objection which is too difficult (and too long) to discuss here is that truth, necessity, knowledge, and so on, apply to objects of different sorts, like sentences, propositions, utterances, etc.

References

FAGIN, R., HALPERN, J., MOSES, Y. and VARDI, M. (2003) *Reasoning about Knowledge*. Cambridge, Mass: MIT.

HALBACH, V. (2006) "How Not to State T-sentences." *Analysis* 66: 276–280.

HALBACH, V. (2008) "On a Side Effect of Solving Fitch's Paradox by Typing Knowledge." *Analysis* 68(2): 114–120.

HALBACH, V. and WELCH, P. (2009) "Necessities and Necessary Truths: A Prolegomenon to the Use of Modal Logic in the Analysis of Intensional Notions." *Mind* 118(469): 71–100.

HORSTEN, L. (2009) "Levity." *Mind* 118(471): 555–581.

HORSTEN, L. and LEITGEB, H. (2001) "No Future." *Journal of Philosophical Logic* 30: 259–265.

KRIPKE, S. (1975) "Outline of a Theory of Truth." *Journal of Philosophy* 72(19): 690–716.

LINSKY, L. (2009) "Logical Types in Some Arguments about Knowability and Belief." In J. Salerno (ed.), *New Essays on the Knowability Paradox*, pp. 163–179. New York: Oxford University Press.

MONTAGUE, R. (1963) "Syntactical Treatments of Modality with Corollaries on Reflection Principles and Finite Axiomatizability." *Acta Philosophica Fennica* 16: 153–167.

PASEAU, A. (2008) "Fitch's Argument and Typing Knowledge." *Notre Dame Journal of Formal Logic* 49(2): 155–176.

PASEAU, A. (2009) "How to Type: Reply to Halbach." *Analysis* 69(2): 280–286.

ROSENBLATT, L. (2014) "The Knowability Argument and the Syntactic Type-theoretic Approach." *Theoria. An International Journal for Theory, History and Foundations of Science*, forthcoming.

SALERNO, J. (ed.) (2009) *New Essays on the Knowability Paradox*. New York: Oxford University Press.

SMORYNSKI, C. (2004) "Modal Logic and Self-reference." In D. Gabbay and F. Guethner (eds), *Handbook of Philosophical Logic* (2nd edition), Vol. 11, pp. 1–55. Dordrecht: Kluwer.

STERN, J. and FISCHER, M. (2013) "Paradoxes of Interaction?" Manuscript.