

# The Knowability Argument and the Syntactic Type-Theoretic Approach\*

Lucas ROSENBLATT

Received: 04.02.2013

Final version: 08.10.2013

BIBLID [0495-4548 (2014) 29: 80; pp. 201-221]

DOI: 10.1387/theoria.7225

**ABSTRACT:** Recently, there have been some attempts to block the Knowability Paradox and other modal paradoxes by adopting a type-theoretic framework in which knowledge and necessity are regarded as typed predicates. The main problem with this approach is that when these notions are simultaneously treated as predicates, a new kind of paradox appears. I claim that avoiding this paradox either by weakening the Knowability Principle or by introducing types for both predicates is rather messy and unattractive. I also consider the prospect of using the truth predicate to emulate necessity, knowledge and other modal notions. It turns out that this idea works much better.

**Keywords:** Knowability Argument; Type-theoretic approach; Self-reference; Multimodal paradoxes; Truth.

**RESUMEN:** Recientemente, ha habido intentos por resolver la Paradoja de la Cognoscibilidad y otras paradojas modales por medio de la adopción de un enfoque de tipos en el cual las nociones de conocimiento y necesidad se representan utilizando predicados tipeados. El principal problema con esta propuesta es que cuando estas nociones son tratadas simultáneamente como predicados, una nueva clase de paradoja aparece. En este artículo sostengo que evitar esta paradoja debilitando el Principio de Cognoscibilidad o introduciendo tipos para ambos predicados no es una solución atractiva. También considero una propuesta alternativa, la de utilizar el predicado veritativo para emular las nociones de necesidad, conocimiento y otras nociones modales. Resulta que esta última idea funciona mucho mejor.

**Palabras clave:** El Argumento de la Cognoscibilidad; Enfoque de tipos; Autorreferencia; Paradojas multimodales; Verdad.

## 1. Introduction

Recently some attempts have been made to block the Knowability Argument (sometimes also called the Knowability Paradox) and certain modal paradoxes by adopting a type-theoretic framework in which knowledge and necessity are regarded as typed predicates. The main problem with this approach is that when these notions are simultaneously treated as (typed or untyped) predicates, a new kind of paradox—one that is immune to type restrictions—can be constructed. I claim that avoiding this paradox either by proposing a very weak form of the Knowability Principle or by introducing types for all paradox-generating predicates is rather messy and unattractive. I also consider the prospect of using the truth predicate to emulate other modal notions. It turns out that this idea has several attractive features.

---

\* I am very grateful to two anonymous referees for their comments. I owe special thanks to Eduardo Barrio, Eleonora Cresto, Paulo Faría, Volker Halbach, Federico Pailos, Lavinia Picollo, Diego Tajer and to audiences at the *Third New College Logic Meeting* (Oxford) and *Segundo Congreso Latinoamericano de Filosofía Analítica* (Buenos Aires). This paper would not have been written without the financial support of CONICET.



The paper is structured as follows. In section 2 I offer a version of the Knowability Argument and a sketch of how the type theoretic approach is useful to block not only this argument but also the usual paradoxes of self-reference. In section 3 I introduce a problem for the type-theoretic approach posed by Volker Halbach and I present three possible replies available to the type-theorist. The rest of the paper considers each of these replies. In section 4 I consider the possibility of embracing only a very weak form of the Knowability Principle. In section 5 I evaluate the idea of having types for every modal predicate. It turns out that both options are subject to serious problems. In section 6 I argue for the idea of using truth to emulate modal notions such as knowledge and necessity. In section 7 I build a simplified axiomatic system involving truth, necessity and knowledge to illustrate the sort of theory I have in mind. In section 8 I show how this axiomatic system blocks every self-referential paradox expressible in the system while validating the Knowability Argument. Section 9, the last one, contains some closing remarks.

## 2. The Knowability Argument

In recent years the Knowability Argument<sup>1</sup> (sometimes called the “Knowability Paradox” or the “Fitch-Church Paradox”) has provoked a number of different reactions. Some philosophers think that the argument motivates the rejection of the so-called “Knowability Principle”, according to which every truth is knowable.

$$[\text{KP}] \quad \varphi \rightarrow \Diamond \ulcorner K \urcorner \varphi \urcorner \quad \text{Knowability Principle}^2$$

It is quite simple to show that this principle together with other reasonable principles and rules for knowledge and necessity lead to unpleasant consequences. Assume that the following principles from modal and epistemic logic hold:

$$[\text{K}\wedge\text{-dist}] \quad K \ulcorner \varphi \wedge \psi \urcorner \rightarrow K \ulcorner \varphi \urcorner \wedge K \ulcorner \psi \urcorner \quad \text{Distribution of } K \text{ over } \wedge$$

$$[\text{K-fact}] \quad K \ulcorner \varphi \urcorner \rightarrow \varphi \quad \text{Factivity of Knowledge}$$

$$[\Box\text{-nec}] \quad \text{If } \vdash \varphi, \text{ then } \vdash \Box \ulcorner \varphi \urcorner \quad \text{Necessitation for Necessity}$$

It is clear then that using these principles we can prove  $\Box \ulcorner \neg K \ulcorner \varphi \wedge \neg K \ulcorner \varphi \urcorner \urcorner \urcorner$ , or equivalently,  $\neg \Diamond \ulcorner K \ulcorner \varphi \wedge \neg K \ulcorner \varphi \urcorner \urcorner \urcorner$ . For we have

$$K \ulcorner \varphi \wedge \neg K \ulcorner \varphi \urcorner \urcorner \rightarrow K \ulcorner \varphi \urcorner \wedge K \ulcorner \neg K \ulcorner \varphi \urcorner \urcorner \quad \text{by } [\text{K}\wedge\text{-dist}]$$

$$K \ulcorner \varphi \urcorner \wedge K \ulcorner \neg K \ulcorner \varphi \urcorner \urcorner \rightarrow K \ulcorner \varphi \urcorner \wedge \neg K \ulcorner \varphi \urcorner \quad \text{by } [\text{K-fact}]$$

<sup>1</sup> The reader interested in the Knowability Argument can see Fitch’s classic paper (Fitch 1963). More updated versions of the paradox and some discussions surrounding it can be found in (Salerno 2009).

<sup>2</sup> Because of certain aspects of the discussion that will come up later, possibility  $\Diamond(x)$ , necessity  $\Box(x)$  and knowledge  $K(x)$  are treated as predicates. I will assume that every formula of the language has a name. So for any formula  $\varphi$ , the expression  $\ulcorner \varphi \urcorner$  should be understood as a term that denotes the (code of the) sentence  $\varphi$  (for an appropriate coding of the expressions of the language). Also, the knowledge predicate is usually thought to be implicitly indexed by a time and an agent.  $K \ulcorner \varphi \urcorner$  stands for “someone at some time knows that  $\varphi$ ”. Unless stated otherwise, throughout the paper the agent and the time indexes will be kept fixed and hence will be omitted.

$$K \ulcorner \varphi \wedge \neg K \ulcorner \varphi \urcorner \urcorner \rightarrow K \ulcorner \varphi \urcorner \wedge \neg K \ulcorner \varphi \urcorner \quad \text{by Logic}$$

This means that  $K \ulcorner \varphi \wedge \neg K \ulcorner \varphi \urcorner \urcorner$  leads to a contradiction. But then, classical logic gives us:

$$\neg K \ulcorner \varphi \wedge \neg K \ulcorner \varphi \urcorner \urcorner$$

Applying  $[\Box\text{-nec}]$  we obtain a formula equivalent to

$$\neg \Diamond \ulcorner K \ulcorner \varphi \wedge \neg K \ulcorner \varphi \urcorner \urcorner \urcorner$$

Now notice that a possible instance of [KP] is the following:

$$\varphi \wedge \neg K \ulcorner \varphi \urcorner \rightarrow \Diamond \ulcorner K \ulcorner \varphi \wedge \neg K \ulcorner \varphi \urcorner \urcorner \urcorner.$$

But then logical reasoning gives us

$$\neg(\varphi \wedge \neg K \ulcorner \varphi \urcorner)$$

which is (classically) equivalent to

$$\varphi \rightarrow K \ulcorner \varphi \urcorner^3$$

This last sentence informally asserts that every truth is known. But most people (those who do not subscribe to a very strong form of verificationism) will agree that the Principle of Non-Omniscience is acceptable:

(Non-Omniscience) Some truth is not known.

Therefore, if the other modal rules are sound and classical logic is not questioned, we end up with an inconsistency: every truth is known and some truth is not known.

This argument is usually thought to be problematic for the anti-realist philosopher, since the thesis that every truth is knowable is taken to be a central ingredient of her viewpoint. This has to do with the fact that the principle relates truth to knowability, the idea being that there are no truths beyond the epistemic capacity of the potential agents. As a matter of fact, the proof above is sometimes considered as a direct refutation of various forms of anti-realism.

There is a different reading of the argument available. Some philosophers think that it is a mistake to frame the discussion as one between realists and anti-realists. They claim that the argument is a genuine paradox, since it collapses the reasonable (although perhaps false) philosophical claim that every truth is knowable into the unacceptable claim that every truth is known.<sup>4</sup>

Under this reading, it becomes crucial to block the argument. Some philosophers believe that the Knowability Principle can be saved (and the paradox avoided) if some restriction on the set of instances of the Principle is set up. Among the many ways to restrict the Knowability Principle discussed in the literature, I am only interested here in the possibility of adopting a type-theoretic restriction. The advantage of adopting a

<sup>3</sup> In this derivation a couple of simplifications were made. First, instead of using quantification over propositions or over objects I used schematic versions of the relevant principles. Second, I was being very sloppy with the notation. For example, in  $K \ulcorner \neg \varphi \urcorner$  the symbol  $\neg$  is not the negation symbol but a function that applied to the code of  $\varphi$  gives the code of its negation.

<sup>4</sup> See (Kvandig 2008) for a defense of this reading. Some (convincing) arguments against it can be found in (Jenkins 2009).

solution of this sort is evident. Since other well-known paradoxes can be avoided using type restrictions, the type-theoretic framework can provide a unified answer to many (maybe all) different paradoxes, assuming that the Knowability Argument indeed constitutes a paradox.<sup>5</sup>

To start with a familiar example, take the truth predicate  $T$ . There are different ways of implementing type restrictions within a theory of truth. One possibility is to syntactically ban all sentences of the form  $T \ulcorner \varphi \urcorner$  if  $\varphi$  contains an occurrence of  $T$ . Another possibility is to restrict the proof-theoretic principles in which  $T$  is involved. For example, if we have a base theory  $\mathbb{T}$  capable of self-reference and  $T$  is added to the language  $\mathcal{L}_{\mathbb{T}}$  of  $\mathbb{T}$ , instead of accepting the full unrestricted  $T$ -schema

$$(T\text{-schema}) \quad T \ulcorner \varphi \urcorner \leftrightarrow \varphi$$

we only accept those instances of the  $T$ -schema in which  $\varphi$  is a formula of the language  $\mathcal{L}_{\mathbb{T}}$ . In this paper I will consider only typed theories of this sort, although nothing crucial hangs on this.

A more sophisticated version of the proof-theoretic way of typing is the following. We start by introducing infinitely many truth predicates  $T_1, T_2, T_3, \dots$  to the language of the base theory  $\mathbb{T}$ . The idea is that  $\mathcal{L}_{\mathbb{T}} \cup \{T_1\}$  is  $\mathcal{L}_{\mathbb{T}T_1}$ ,  $\mathcal{L}_{\mathbb{T}} \cup \{T_1, T_2\}$  is  $\mathcal{L}_{\mathbb{T}T_1T_2}$ , and so on. For each new language  $\mathcal{L}_{\mathbb{T}T_1 \dots T_m}$  we have a new restricted version of the  $T$ -schema, so that for each  $n$  such that  $m \geq n > 0$ ,  $T_n \ulcorner \varphi \urcorner \leftrightarrow \varphi$  holds if and only if  $\varphi$  contains no occurrences of  $T_j$ , for  $m \geq j \geq n$ . The resulting language is  $\mathcal{L}_{\mathbb{T}^{\infty}}$ , that is,  $\mathcal{L}_{\mathbb{T}} \cup \{T_1, T_2, T_3, \dots\}$ .<sup>6</sup> The resulting theory has, for each  $n > 0$ , a restricted  $T$ -schema for  $T_n$ . More formally, for each  $n > 0$  the following holds:

$$(\text{Restricted } T\text{-Schema}) \quad T_n \ulcorner \varphi \urcorner \leftrightarrow \varphi \quad \text{where } \varphi \text{ does not contain occurrences of } T_m \text{ for } m \geq n.$$

It will be useful to have a definition of the type of a formula. I will say that (for  $m > 0$ ) the truth-type  $\tau(\varphi)$  of a formula  $\varphi$  is  $m$ , more formally,  $\tau(\varphi) = m$ , if  $\varphi$  contains an occurrence of  $T_m$  and for every  $n > m$ ,  $\varphi$  does not contain any occurrence of  $T_n$ , otherwise,  $\tau(\varphi) = 0$ . Of course, for every formula  $\varphi \in \mathcal{L}_{\mathbb{T}}$ ,  $\tau(\varphi) = 0$ .

It should be clear that a liar sentence, for example, is no longer an instance of the Restricted  $T$ -schema because in

$$T_m \ulcorner \varphi \urcorner \leftrightarrow \varphi \quad \text{where } \tau(\varphi) < m$$

we cannot replace  $\varphi$  by the formula  $\neg T_m \ulcorner A \urcorner$  without violating a type restriction. We can replace it by  $\neg T_n \ulcorner A \urcorner$  if  $m > n$ , thus obtaining

$$T_m \ulcorner \neg T_n \ulcorner A \urcorner \urcorner \leftrightarrow \neg T_n \ulcorner A \urcorner.$$

But this will not generate a contradiction.

<sup>5</sup> Curiously, this kind of solution was anticipated by Alonzo Church in an anonymous referee report to Fitch's article where the Knowability Argument first appeared. The reader can see (Salerno 2009) for the details.

<sup>6</sup> Of course, the hierarchy of truth predicates could continue at transfinite levels, but to simplify things I will not consider this possibility here.

The similarities between the concept of truth and other concepts like necessity and knowledge is usually overlooked. The reaction of the philosophical community towards the Liar paradox is very different from its reaction towards modal Liar-like paradoxes such as the Knower. Regarding the first one, many different solutions have been proposed: disallowing self-reference, modifying classical logic, using revision sequences, among other things. Concerning the others, the traditional reaction is to treat the corresponding notions as operators instead of treating them as predicates. This difference in attitude might seem strange since it is perfectly possible to use the other strategies (banning self-reference, dropping classical logic, etc.) to solve the modal paradoxes. So the question we must ask becomes clear: why is it so common (nowadays) to treat modalities as operators instead of treating them as predicates? The answer can probably be found in (Montague 1963). Montague proves a version of the following theorem:

(*Theorem 1*) Let  $\mathbb{T}$  be a theory that extends Robinson's Arithmetic  $Q^7$  and let  $P(x)$  be a monadic predicate of  $\mathcal{L}_{\mathbb{T}}$  (the language in which the theory  $\mathbb{T}$  is formulated). If  $P(x)$  satisfies (P-fact) and (P-nec), then  $\mathbb{T}$  is inconsistent.

(P-fact)  $P \ulcorner \varphi \urcorner \rightarrow \varphi$

(P-nec) If  $\vdash_{\mathbb{T}} \varphi$ , then  $\vdash_{\mathbb{T}} P \ulcorner \varphi \urcorner$

The proof is very simple. Informally, it is enough to consider the following instance of the Diagonal Lemma:  $A \leftrightarrow \neg P \ulcorner A \urcorner$ . Assume that  $P \ulcorner A \urcorner$  holds. By (P-fact), we obtain  $A$  and using the biconditional above we infer  $\neg P \ulcorner A \urcorner$ . By *reductio*, we have  $\neg P \ulcorner A \urcorner$ , which is equivalent to  $A$ . If we apply (P-nec) to  $A$  we obtain  $P \ulcorner A \urcorner$ . So we have a contradiction. This proof is quite general. It applies to every predicate satisfying Factivity and Necessitation. Two prominent examples are the knowledge and the necessity predicates.

For a long time the philosophical community took this result as conclusive evidence that modalities must be treated as operators and not as predicates. For instance, Montague claims 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" (1963, 294).

The comparison with the concept of truth motivates the following question: why can't we avoid modal paradoxes by modifying some of the principles that govern the modal predicates? One traditional reaction to the Liar (and related semantic paradoxes like the Curry paradox) is to restrict self-reference by adopting a type-theoretic framework. This approach is very effective when it comes to blocking the usual semantic inconsistencies.<sup>8</sup>

---

<sup>7</sup> Or any theory in which the Diagonal Lemma holds.

<sup>8</sup> Of course, there are reasons to question the adequacy of such an approach. These reasons are usually related either to the unnaturalness of the types or to the expressive and proof-theoretic weakness of

So nothing prevents us from doing the same with the modal notions understood as predicates. Let  $P_1, P_2, P_3, \dots$  be modal predicates such that a formula  $\varphi$  has P-type  $m$ , that is,  $\pi(\varphi) = m$  if it has an occurrence of  $P_m$  and for any  $n > m$ , it has no occurrences of  $P_n$ . Otherwise,  $\pi(\varphi) = 0$ . One possible way to implement the type restrictions for formulas containing  $P_i$  is to regard as legitimate only certain instances of Factivity and Necessitation for  $P_i$ :

(Restricted P-fact)  $P_m \ulcorner \varphi \urcorner \rightarrow \varphi$  where  $\pi(\varphi) < m$ .

(Restricted P-nec) If  $\vdash_{\mathbb{T}} \varphi$ , then  $\vdash_{\mathbb{T}} P_m \ulcorner \varphi \urcorner$  where  $\pi(\varphi) < m$ .

Under these conditions, Montague's paradox no longer arises.

Also, the Knowability Argument is no longer available. If we try to mimic the proof presented at the beginning, in virtue of the type restrictions for knowledge we would only have something like

$$K_m \ulcorner \varphi \wedge \neg K_n \ulcorner \varphi \urcorner \urcorner \rightarrow (K_m \ulcorner \varphi \urcorner \wedge K_m \ulcorner \neg K_n \ulcorner \varphi \urcorner \urcorner)$$

where  $\kappa(\varphi) < n < m$  and where  $\kappa(\varphi)$  stands for the knowledge-type of the formula  $\varphi$ . From this, we can infer

$$K_m \ulcorner \varphi \wedge \neg K_n \ulcorner \varphi \urcorner \urcorner \rightarrow (K_m \ulcorner \varphi \urcorner \wedge \neg K_n \ulcorner \varphi \urcorner)$$

But the consequent is far from being a contradiction. We can obtain a contradiction if the following principle holds:

$$K_m \ulcorner \varphi \urcorner \rightarrow K_n \ulcorner \varphi \urcorner \quad \text{for every } m, n \text{ such that } \kappa(\varphi) < n < m.$$

But those who favor a typed theory of knowledge have reasons to reject it. Hence the Knowability Argument is blocked.

### 3. A new paradox

One recent defense of this approach has been provided by Alexander Paseau (cf. 2008 and 2009). He claims that the following principle fails if  $P(x)$  represents knowledge:

(Minimality P) For any  $n > \pi(\varphi)$ , if  $P_n \ulcorner \varphi \urcorner$ , then  $P_{\pi(\varphi)+1} \ulcorner \varphi \urcorner$ .

Paseau provides an independent conceptual motivation for thinking that this principle fails. 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: "the typing is one of epistemic access rather than (just) content" (2008, 163). For example, if  $A$  is a sentence and  $\kappa(A)=0$ , then, if the state of affairs depicted by  $A$  is known by an agent through per-

---

the resulting theory. Below I will explore some additional complications that arise when the knowledge predicate is typed.

<sup>9</sup> If we employ a typed language where only formulas of the form  $K_{\kappa(\varphi)+1} \ulcorner \varphi \urcorner$  are permitted, the paradox will be blocked as well because from the assumption  $K_{\kappa(\varphi)+2} \ulcorner \varphi \wedge \neg K_{\kappa(\varphi)+1} \ulcorner \varphi \urcorner \urcorner$  we cannot infer  $K_{\kappa(\varphi)+2} \ulcorner \varphi \urcorner \wedge K_{\kappa(\varphi)+2} \ulcorner \neg K_{\kappa(\varphi)+1} \ulcorner \varphi \urcorner \urcorner$  since the first conjunct is not well formed. A careful examination of this strategy is beyond the scope of this paper. Let me just say that I have qualms about this way of implementing the type restrictions. In particular, the resulting language is expressively weak and reasonable instances of Distribution, like the one above, are not validated.

ceptual means, we might say that the agent has type 1 knowledge of A, that is, the sentence  $K_1 \ulcorner A \urcorner$  is true. However, not every state of affairs depicted by a type 0 sentence can be known in this way. Think of a sentence B such that we do not have (and cannot have) perceptual access to the state of affairs depicted by it and such that B is only known indirectly through our knowledge of other states of affairs depicted the sentences C and D. Then we could say that we have type 2 (or maybe type 3) knowledge of B, that is, the sentence  $K_2 \ulcorner B \urcorner$  would be true but the sentence  $K_1 \ulcorner B \urcorner$  would be false. The following example is provided by Paseau: “I know<sub>2</sub> that Giggs scored the winning penalty in the 2008 Champions League final if I infer it from the facts that you know<sub>1</sub> that it was either Giggs or Ronaldo, and that Ronaldo missed his penalty” (2009, 284). Also Linsky subscribes to a similar line of reasoning: “We need not in general accept the principle  $K_2 p \rightarrow K_1 p$ . (...) 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” (2009, 172).

In Paseau’s proposal  $K$  and  $\Box$  are treated as predicates. This has some advantages. First, treating the modalities as predicates has the virtue of increasing the expressive capacity of the language. We can assert non-schematic generalizations such as

$$\forall x \forall y (K(x) \ \& \ K(x \rightarrow y) \rightarrow K(y)) \text{ and} \\ \exists x K(x),^{10}$$

to give a couple of examples, that would only be expressible by the schemas

$$K(\varphi) \ \& \ K(\varphi \rightarrow \psi) \rightarrow K(\psi) \text{ and} \\ K\varphi \text{ for some } \varphi$$

if knowledge were an operator (unless we were willing to admit quantification over propositions).<sup>11</sup> Second, we can construct self-referential sentences in the well-studied framework of arithmetic using the Diagonal Lemma. This is a virtue at least to the extent that we expect our theories to prove all sorts of truths, even truths about their own sentences.<sup>12</sup>

Naturally, the presence of the Diagonal Lemma implies that the risk of generating an inconsistency is very high. In particular, it is not difficult to think of problematic instances of the Lemma in which two modalities are combined. For example, consider the sentence ‘This sentence is not knowable’ or the sentence ‘If this sentence is know-

<sup>10</sup> Again, I am being sloppy with the notation. The possible values of  $x$  are restricted to (names of) sentences in both cases, and the conditional occurring in the scope of the knowledge predicate is a function rather than the actual connective.

<sup>11</sup> But adopting quantification over propositions is all but innocent. For one thing, it amounts to accepting a form of second-order quantification. For another, we would be stuck (without any real need) with two sort of quantifiers: the usual first-order objectual ones, and the new propositional quantifiers, which are usually substitutionally interpreted.

<sup>12</sup> Paseau also offers a very sophisticated semantic framework showing that the typed Knowability Principle is valid while the claim that every truth is known is invalid. This shows that the Knowability Argument is blocked in his framework.

able, then  $\perp$ , which are modal-epistemic versions of the Liar and the Curry sentence, respectively. Halbach (2008) has shown that because of sentences like these, it is not enough to type the language in the way explained above:

(Theorem 2) Let  $\mathbb{T}$  be a theory that extends  $\mathcal{Q}$  and let  $\Box_m(x)$  and  $K_n(x)$  be predicates of  $\mathcal{L}_{\mathbb{T}}$ . If  $\Box_m(x)$  and  $K_n(x)$  satisfy (Restricted K-fact), (Restricted  $\Box$ -nec) and (Restricted KP), then  $\mathbb{T}$  is inconsistent.

(Restricted K-fact)	$K_n \ulcorner \varphi \urcorner \rightarrow \varphi$	where $\kappa(\varphi) < n$ .
(Restricted $\Box$ -nec)	If $\vdash_{\mathbb{T}} \varphi$ , then $\vdash_{\mathbb{T}} \Box_m \ulcorner \varphi \urcorner$	where $\eta(\varphi) < m$ <sup>13</sup>
(Restricted KP)	$\varphi \rightarrow \Diamond_m \ulcorner K_n \ulcorner \varphi \urcorner \urcorner$	where $\kappa(\varphi) < n$ .

Proof: By the Diagonal Lemma, which is provable  $\mathbb{T}$ , we know that  $\vdash_{\mathbb{T}} A \leftrightarrow \Box_m \ulcorner \neg K_n \ulcorner A \urcorner \urcorner$ .<sup>14</sup> Since  $\kappa(A) < n$ , by (Restricted KP) we have  $\vdash_{\mathbb{T}} A \rightarrow \Diamond_m \ulcorner K_n \ulcorner A \urcorner \urcorner$ . So logic alone gives us  $\vdash_{\mathbb{T}} \neg A$ . Since we also have (Restricted K-fact), we can obtain  $\vdash_{\mathbb{T}} \neg K_n \ulcorner A \urcorner$ , given again that  $\kappa(A) < n$ . But using (Restricted  $\Box$ -nec) together with the fact that  $\eta(\neg K_n \ulcorner A \urcorner) < m$ , we get  $\vdash_{\mathbb{T}} \Box_m \ulcorner \neg K_n \ulcorner A \urcorner \urcorner$ , which yields  $\vdash_{\mathbb{T}} A$ . So  $\mathbb{T}$  is inconsistent.

As we have seen, the Knowability Principle is highly controversial because it leads to an inconsistency. The surprising aspect of Halbach's proof is that a contradiction is obtainable even from a type-restricted version of the Knowability Principle (and other restricted modal principles). So Halbach's proof is an argument against the idea of blocking the Knowability Argument by appealing to types.

There are a few details in Halbach's proof that deserve mentioning. One thing is that the knowledge-types ( $\kappa$ ) and the possibility-types ( $\eta$ ) do not interact. Of course, it would be too extreme to ban all interactions of knowledge with possibility, since in order to express the Knowability Principle the type-theorist needs formulas where these two predicates are combined. Also, there is *prima facie* nothing wrong with sentences of the form  $\Diamond_m \ulcorner K_n \ulcorner \varphi \urcorner \urcorner$  where the possibility-type  $m$  is smaller than the knowledge-type  $n$ . The rules above do not ban constructions of this sort.<sup>15</sup> A second noticeable feature of the proof is that it uses the fact that (Restricted KP) imposes no restriction concerning the appearance of  $\Diamond_m$  (and  $\Box_m$ ) in  $\varphi$ . It is only asked that  $\varphi$  contains no occurrences of  $K_n$ .

The type-theorist has, as far as I can see, at least three ways to avoid Halbach's objection without losing the expressive power gained by treating knowledge and necessity as predicates:

<sup>13</sup>  $\eta(\varphi)$  stands for the necessity(possibility)-type of the formula  $\varphi$ .

<sup>14</sup> The fact that  $\vdash_{\mathbb{T}} A \leftrightarrow \Box_m \ulcorner \neg K_n \ulcorner A \urcorner \urcorner$  means that  $A$  is the sentence  $\exists x(x = \ulcorner A \urcorner \wedge \Box_m \ulcorner K_n \ulcorner x \urcorner \urcorner)$ . So  $\kappa(\neg K_n \ulcorner A \urcorner) = n$  and  $\eta(A) = m$ , but since the predicate  $K_n$  occurs in  $A$  only as part of a name,  $\kappa(A) = \eta(\neg K_n \ulcorner A \urcorner) = 0$ .

<sup>15</sup> Although it could be argued that they should. I will deal with this objection later.

- Applying a stronger type restriction on the Knowability Principle.
- Introducing cross-hierarchy type restrictions.
- Using truth to emulate the knowledge and necessity.

I will argue that only the third option is attractive. I will also argue that this has two consequences. First, it implies that there are no good reasons to type knowledge and necessity, and second, that either the Knowability Argument is sound or it must be blocked without resorting to type restrictions.

4. *Weakening the Knowability Principle*

Paseau claims that one way to avoid Halbach’s proof is to further weaken the Knowability Principle. The idea is to adopt the following weak version of Knowability:

$$\text{(Restricted KP*)} \quad \varphi \rightarrow \hat{\diamond}_m \ulcorner K_n \ulcorner \varphi \urcorner \urcorner \quad \text{for some } n \geq \kappa(\varphi)^{16}$$

The difference with the previous version of Knowability is that something, say  $A$ , can be knowable at type  $\kappa(A)+n$  without being knowable at type  $\kappa(A)+1$ , where  $n > 1$ .

I claim that the restriction imposed on the Knowability Principle is problematic and, more importantly, not strong enough to block certain paradoxes. First of all, (Restricted KP\*) quantifies (apparently in an unrestricted way) over types. So the type-theorist needs to slip in type quantifiers into the language. Otherwise, the Knowability Principle is not even expressible. It is clear, though, that if quantification over the type hierarchy is admissible, then the language is capable of making problematic general claims such as “For every knowledge-type  $x$  greater than  $\kappa(A)$ ,  $K_x \ulcorner A \urcorner$ ”, “ $A \wedge$  for no knowledge-type  $x$  greater than  $\kappa(A)$ ,  $K_x \ulcorner A \urcorner$ ”, and “For some knowledge-type  $x$  greater than  $\kappa(A)$ ,  $K_x \ulcorner A \wedge B \urcorner$ ”.

But what is the knowledge-type of these assertions? And, more prominently, what is the knowledge-type of the (Restricted KP\*) itself? The type-theorist could argue that not every formula has a knowledge-type. A possible way to implement this idea is to define the type function  $\kappa$  as a partial function, thus allowing for quantifications over knowledge-types while avoiding the problem of assigning a type to those formulas that unboundedly quantify over types.

However, if quantification over knowledge-types is admissible, a contradiction is still obtainable with (Restricted KP\*) provided that the formulas in which the knowledge-type quantifiers appear have no specific knowledge-type and that the restricted Factivity principle introduced above is applicable to these formulas. In other words, we have the following result

---

<sup>16</sup> Sometimes Paseau uses operators for knowledge and possibility, and sometimes he uses operators. To keep things uniform with the rest of the paper, I will use typed predicates for both possibility (necessity) and knowledge.

(Theorem 3) Let  $\mathbb{T}$  be a theory that extends  $\mathcal{Q}$  and let  $\Box_m(x)$  and  $K_n(x)$  be predicates of  $\mathcal{L}_{\mathbb{T}}$ . Assume that  $\mathcal{L}_{\mathbb{T}}$  can quantify over types, i.e.,  $\mathcal{L}_{\mathbb{T}}$  has expressions of the form “for every/some knowledge type  $x$ ” and “for every/some necessity type  $x$ ”. If  $\Box_m(x)$  y  $K_n(x)$  satisfy (Restricted K-fact\*), (Restricted  $\Box$ -nec\*) and (Restricted KP\*), then  $\mathbb{T}$  is inconsistent.

(Restricted  $\Box$ -nec\*) If  $\vdash_{\mathbb{T}} \varphi$ , then  $\vdash_{\mathbb{T}} \Box_m^{\ulcorner} \varphi \urcorner$  where  $\eta(\varphi) < m$  or  $\varphi$  has no knowledge-type.

(Restricted K-fact\*)  $K_m^{\ulcorner} \varphi \urcorner \rightarrow \varphi$  where  $\kappa(\varphi) < m$  or  $\varphi$  has no knowledge-type.<sup>17</sup>

Proof: The idea behind the proof is to run a usual diagonal argument using the sentence that says of itself that it is unknowable at every knowledge-type. So if  $\mathbb{T}$  proves the Diagonal Lemma, then  $\vdash_{\mathbb{T}} A \leftrightarrow$  for every  $x > \kappa(A)$ ,  $\neg \Diamond_m^{\ulcorner} K_x^{\ulcorner} A \urcorner \urcorner$ . Also, by (Restricted KP\*), we have  $\vdash_{\mathbb{T}}$  For some  $x > \kappa(A)$ ,  $A \rightarrow \Diamond_m^{\ulcorner} K_x^{\ulcorner} A \urcorner \urcorner$ . From these two facts, logical reasoning within  $\mathbb{T}$  gives us  $\vdash_{\mathbb{T}} \neg A$ . By (Restricted K-fact\*) we know that  $\vdash_{\mathbb{T}}$  For every  $x > \kappa(A)$ ,  $K_x^{\ulcorner} A \urcorner \rightarrow A$ . This means that we can infer  $\vdash_{\mathbb{T}} \neg K_{\kappa(A)+n}^{\ulcorner} A \urcorner$ . Using (Restricted  $\Box$ -nec\*), we can obtain  $\vdash_{\mathbb{T}} \Box_m^{\ulcorner} \neg K_{\kappa(A)+n}^{\ulcorner} A \urcorner \urcorner$ , which is equivalent to the right hand side of the biconditional above. So we also have  $\vdash_{\mathbb{T}} A$ . Hence,  $\mathbb{T}$  is inconsistent.<sup>18</sup>

But is the quantification over types something the type-theorist would accept? Well, remember that in order to express (Restricted KP\*) (as presented by Paseau) we must quantify over types<sup>19</sup>. The type-theorist could reply that it is possible to dispose of the quantifier if (Restricted KP\*) is understood as a schema, i.e., if *some* instances of the Knowability Principle are taken as axioms. However, this move is only available for the first version of the Knowability Principle, not for (Restricted KP\*). Recall that this principle involves an *existential* quantification over types. We do not know which instances of  $n$  should be taken as axioms.

<sup>17</sup> It would be pointless for the advocate of the Knowability Principle to question the idea that the knowledge predicate applies to formulas of the form “For every knowledge-type  $x$ , ...” or “For some knowledge-type  $x$ , ...” once they have accepted them as part of the language. If no knowledge predicate were applicable to this sort of sentences, there would automatically be unknowable sentences in the language. The advocate of the Knowability Principle certainly does not want that.

<sup>18</sup> Notice that in order to obtain an unpleasant result a diagonal argument is not really necessary here. We could use a Fitch-style reasoning giving us the conclusion that if every truth is *knowable* at some type, then every truth is *known* at some type, a result that seems quite unreasonable. The argument is briefly sketched in (Hart 2009).

<sup>19</sup> A similar point is made by (Carrara and Fassio 2011, 187-88) about the Non-Omniscience Principle. Intuitively, the occurrence of the knowledge predicate (operator) in this claim seems to be type-unrestricted. So in order to express it in a type-theoretic framework, we would need to quantify over types:  $\varphi \wedge$  for some  $x > \kappa(\varphi)$ ,  $\neg K_x^{\ulcorner} \varphi \urcorner$ .

To deal with this issue, Paseau (2009, 285) tentatively proposes to introduce a knowability function  $k$ , which applies to a sentence and outputs the type at which the sentence is first knowable. He suggests that the type-theorist can endorse the schema  $\varphi \rightarrow \Diamond K_{k(\varphi)}\varphi$  and notes that we cannot resurrect Halbach's diagonal argument using  $\gamma \leftrightarrow \neg\Diamond^{\ulcorner} K_{k(\gamma)}^{\urcorner} \gamma^{\urcorner}$ . For to run the diagonal argument, we first fix an  $n$  and then diagonalize on the predicate  $\neg\Diamond^{\ulcorner} K_n(x)^{\urcorner}$  obtaining  $\gamma \leftrightarrow \neg\Diamond^{\ulcorner} K_n^{\urcorner} \gamma^{\urcorner}$  for some sentence  $\gamma$ . But this  $\gamma$  is such that  $k(\gamma) \neq n$ .

However, it is unclear to me whether this version of the Knowability Principle is an improvement over the existential version. The knowledge-types are supposed to represent the way in which a particular proposition has come to be known by an agent, so it seems that to calculate the output of the function for a certain input some empirical information is needed. Hence, we only find out what the admissible instances of  $\varphi \rightarrow \Diamond K_{k(\varphi)}\varphi$  are by doing empirical research. This option does not seem to be very attractive.

Moreover, even if this version of the Knowability Principle could be defended somehow, it would be strange for the supporter of the *syntactic* approach to reject the use of quantifications given that she has already presented reasons for preferring them over schemata. One of the main advantages of treating knowledge as a predicate instead of treating it as an operator is that we can replace schemas with quantifications. So either quantification over types is admissible and the previous proofs are legitimate, or quantification over types is not admissible and the Knowability Principle cannot be adequately expressed.

##### 5. Generalizing Halbach's argument and applying cross-hierarchy type restrictions

At least *prima facie* Halbach's proof only has an impact on philosophers who are keen on the Knowability Principle. Thus, a natural reaction for someone who advocates the project of representing modalities as predicates is simply to reject this principle (even in its restricted version). So, when things are put this way, Halbach's proof can be seen as just one more reason to reject the claim that every truth is knowable. However, it can be shown (Paseau 2009) that a different proof can be constructed without the Knowability Principle, even in its weak version.

(Theorem 4) Let  $\mathbb{T}$  be a theory that extends Q and let  $\Box_m(x)$  and  $K_n(x)$  be predicates of  $\mathcal{L}_{\mathbb{T}}$ . If (Restricted K-fact), (Restricted  $\Box$ -fact), (Restricted  $\Box$ -nec) and (Restricted K-nec) hold, then  $\mathbb{T}$  is inconsistent.

Proof: Since  $\mathbb{T}$  proves the Diagonal Lemma we have  $\vdash_{\mathbb{T}} A \leftrightarrow \Box_m^{\ulcorner} \neg K_n^{\urcorner} A^{\urcorner}$ . But since (Restricted K-fact) and (Restricted  $\Box$ -fact) both hold, we can infer that  $\vdash_{\mathbb{T}} K_n^{\ulcorner} A^{\urcorner} \rightarrow \neg K_n^{\ulcorner} A^{\urcorner}$ . This in turn gives us  $\vdash_{\mathbb{T}} \neg K_n^{\ulcorner} A^{\urcorner}$ . Using (Restricted  $\Box$ -nec) we obtain  $\vdash_{\mathbb{T}} \Box_m^{\ulcorner} \neg K_n^{\urcorner} A^{\urcorner}$ . By the biconditional above and (Restricted K-nec), we can prove  $\vdash_{\mathbb{T}} K_n^{\ulcorner} A^{\urcorner}$ . Hence  $\mathbb{T}$  is inconsistent.

As a matter of fact, a more general version of this holds. The inconsistency might be viewed as emerging from the interaction of *any* two (or more) predicates satisfying restricted versions of Factivity and Necessitation. More rigorously, if  $\mathbb{T}$  is a theory that extends  $\mathbb{Q}$  and  $P_1(x), \dots, P_n(x)$  are unary predicates satisfying Restricted Factivity and Restricted Necessitation, then  $\mathbb{T}$  is inconsistent. The proof is similar to the previous one. The idea is just to apply Necessitation and Factivity as much times as it takes to the following multimodal instance of the Diagonal Lemma  $A \leftrightarrow \neg P_n \ulcorner \dots \urcorner P_2 \ulcorner P_1 \ulcorner A \urcorner \urcorner \dots \urcorner$ .

However, as we will shortly see, using the existence of multimodal paradoxes to argue that the Knowability Principle is innocent is not the right dialectical move for the type-theorist. Even if she has shown that the previous paradox (the one in *Theorem 4*) makes no use of the Knowability Principle, it is still a paradox that affects her way of conceiving knowledge and necessity. The point is that in order to deal with this paradox, the type-theorist might want to make the necessity(possibility)-types  $\eta$  interact with the knowledge types  $\kappa$ , claiming that it is strange to have different and isolated hierarchies for knowledge and necessity (possibility).

As a matter of fact, Paseau argues that “it is natural to assume [that] there are such restrictions” (2009, 282). 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 *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 necessity is acquired by an agent” or “the way in which something has become necessary”. So, in the case of necessity (possibility), the types, if there are any, should be introduced for different reasons.

In any case, as we said before, it would be too extreme to ban all interactions between predicates. One way of setting up some restrictions without banning all interactions is by linearly ordering the predicates. We sequentially introduce infinitely many predicates  $P_1, P_2, P_3, \dots$  to the language of the base theory  $\mathbb{T}$ . The idea is that  $\mathcal{L}_{\mathbb{T} \cup \{P_1\}}$  is  $\mathcal{L}_{\mathbb{T} P_1}$ ,  $\mathcal{L}_{\mathbb{T} \cup \{P_1, P_2\}}$  is  $\mathcal{L}_{\mathbb{T} P_1 P_2}$ , and so on. For each new language  $\mathcal{L}_{\mathbb{T} P_1 \dots P_n}$  we have new restricted principles for  $P$ . The resulting language is  $\mathcal{L}_{\mathbb{T} P_n}$ , that is,  $\mathcal{L}_{\mathbb{T} \cup \{P_1, P_2, P_3, \dots\}}$ .<sup>20</sup> The resulting theory  $\mathbb{T}_{P_n}$  has, for each  $n > 0$ , restricted versions P-fact and P-necessitation for each  $P_n$ . More formally, for each  $n > 0$ , the following holds:

$$\begin{array}{ll} (P_n\text{-fact}^*) & \vdash_{\mathbb{T}} P_n \ulcorner \varphi \urcorner \rightarrow \varphi & \text{for each } \varphi \in \mathcal{L}_{\mathbb{T} P_1, \dots, P_{n-1}} \\ (P_n\text{-nec}^*) & \text{If } \vdash_{\mathbb{T}} \varphi, \text{ then } \vdash_{\mathbb{T}} P_n \ulcorner \varphi \urcorner & \text{for each } \varphi \in \mathcal{L}_{\mathbb{T} P_1, \dots, P_{n-1}}. \end{array}$$

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

It is straightforward to check that all the proofs given above can be blocked if cross-hierarchy restrictions are set up in this way. Technically this solution works

<sup>20</sup> Again, it is clear that the hierarchy of predicates could continue at transfinite levels.

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 again truth, necessity and knowledge. To make sure that the combination of these notions does not generate paradoxes, we sequentially introduce infinitely many truth predicates, necessity predicates and knowledge predicates. But it is not clear how the different predicates are to be introduced. We could order the predicates like this  $T_1, \Box_2, K_3, T_4, \Box_5, K_6, \dots$  or we could order them like this  $\Box_1, K_2, T_3, \Box_4, K_5, T_6, \dots$  or we could order them like this  $K_1, T_2, \Box_3, K_4, T_5, \Box_6, \dots$  or  $\dots$ . We could even have three sequences,  $T_1, T_2, T_3, \dots$ ,  $\Box_1, \Box_2, \Box_3, \dots$ ,  $K_1, K_2, K_3, \dots$  and still other less natural orders could be chosen. There is no good reason to prefer one order over another. In other words, there are infinitely many ways of linearly ordering the predicates, but they are all equally unmotivated and artificial.

The previous remarks are only tentative. As expected, I do not have a conclusive argument against all possible ways in which necessity, knowledge and other notions could be simultaneously typed<sup>21</sup>. But it is important to point out what the dialectical situation is concerning the introduction of types when several notions are involved. Since the most natural ways of typing are not available, it seems that the burden of the proof is with the type theorist.

#### 6. Using truth to emulate knowledge and necessity

As we said at the beginning, type-theoretic approaches are tempting because they provide a unified solution to many paradoxes. However, it is not always a good thing to have a unified response to very different kinds of phenomena. It is usually acknowledged that whenever we face the same kind of paradox, we should provide the same kind of solution. But are the Liar paradox, Montague's paradox, the self-referential multimodal paradoxes and so on, of the same kind as the Knowability Paradox (if indeed the Knowability Argument is a paradox)? There are a number of issues that need to be addressed before answering this question. For example, what does it mean to be of the same kind, what is the level of abstraction we are dealing with, etc. In any case, what I want to claim here is just that the best solution the type-theorist has to deal with multimodal paradoxes leaves the Knowability Argument untouched, and that that is as it should be, since it is reasonable to say that the Knowability Argument is not a genuine paradox, or at least that it is not of the same kind as the Liar and other self-referential paradoxes.<sup>22</sup>

<sup>21</sup> There might be a reasonable way of *partially* ordering the predicates. For example, it will work to have predicates  $T_n, \Box_n$  and  $K_n$ , with axioms and rules ( $T_n$ -schema\*\*), ( $\Box_n$ -fact\*\*), ( $K_n$ -fact\*\*), ( $\Box_n$ -nec\*\*), ( $K_n$ -nec\*\*), and so on, such that for instance ( $\Box_n$ -fact\*\*) says that  $\Box_n \varphi \rightarrow \varphi$  is an axiom if  $\varphi$  does not contain occurrences of  $T_m, \Box_m$  or  $K_m$  for  $m \geq n$ . However, it remains unclear what the conceptual motivation is for this sort of typing. For a detailed discussion see (Pailos and Rosenblatt 2013).

<sup>22</sup> Although see (Priest 2009) for an argument to the effect that the Knowability Paradox (just like Russell's paradox, the Liar paradox, and so on) fits his Inclosure schema.

The sort of approach that I am about to put forward is not really new. Some philosophers have explicitly or implicitly endorsed it.<sup>23</sup> My point is merely to stress that the approach provides the type-theorist with the tools she needs to address multimodal paradoxes while not messing with the Knowability Argument. The idea is that instead of representing a certain notion by a predicate  $P$ , we represent it by  $OT$ , where  $O$  is the operator that corresponds to the predicate  $P$  and  $T$  is the truth predicate. I do not claim that  $OT$  expresses exactly the same concept that the operator  $O$  is supposed to express. It is enough for what I want that  $OT$  behaves similarly enough to the modality expressed by the operator  $O$ , and that it avoids the same problems (i.e. paradoxes). In particular, I do not want to subscribe to the idea that  $OT$  has the same meaning as  $O$ . So, in this proposal truth is treated as a predicate but all the other notions are treated as operators (as is usually done).<sup>24</sup>

If this could be generalized, we would have for every operator  $O$  a corresponding predicate of the form  $OT \ulcorner \varphi \urcorner$ . In particular, if  $O$  is  $\Box$  we can think of the predicate  $\Box T(x)$  as capturing the concept of necessity and if  $O$  is  $K$  we can think of the predicate  $KT(x)$  as capturing the concept of knowledge. One might think it is unnatural to replace occurrences of  $P$ , whatever that is, by occurrences of the complex expression  $OT$ . But natural languages allow for paraphrases of this kind. For instance, if  $P$  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.

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  $OT$  are properly understood<sup>25</sup>. For example, the claim that every arithmetical truth is necessary can be expressed by means of a predicate of the form  $\Box T(x)$ , and becomes the claim that every arithmetical truth is necessarily true.

What about multimodal paradoxes? Since the language includes the truth predicate, we need principles for it. Now, whatever the best theory of truth is, by Tarski’s Theorem we know that it will not have the full  $T$ -schema.<sup>26</sup> But in order to infer a contra-

---

<sup>23</sup> See (Halbach and Welch 2009) and (Horsten 2009).

<sup>24</sup> I am not completely sure that the ‘all’ in the previous sentence should be understood unrestrictedly. For example, maybe the notions of (non-idealized) belief and logical validity could be treated as predicates (together with the truth predicate) without generating an inconsistency. It all depends, of course, on the principles governing the corresponding notion.

<sup>25</sup> This idea was sketched by Kripke in his classic paper on truth (Kripke 1975). A similar proposal has been advanced more recently in (Halbach and Welch 2009) at least regarding truth and necessity. However there are at least two differences. First, their aim is different from mine. They are not immediately concerned with multimodal paradoxes but with the possibility of having an expressively rich language in which modal notions are treated as operators. Second, they offer a semantic account in which the meaning of the connectives is given by Strong Kleene logic and the meaning of the truth predicate is given by Kripke’s minimal fixed point construction. I will do something different. I will present a classical axiomatic theory in which there are type restrictions for truth.

<sup>26</sup> Unless, of course, we make substantial revisions on classical logic. However, discussing such theories is well beyond the scope of this paper.

diction from a self-referential multimodal sentence we must have, in addition to the rules for the modal operators, an untyped version of the T-schema. In other words, if the truth predicate only satisfies (Restricted T-schema), then clearly the multimodal paradoxes will no longer be a problem.

### 7. A simple theory

To illustrate how my solution to the multimodal paradoxes works and how the Knowability Argument is preserved, I will present a formal theory in which truth, necessity and knowledge can be represented. Naturally, the solution is meant to be more general, and as a matter of fact other notions could be introduced to the theory. However, for the purpose of illustrating my position, I think the theory I am about to offer is interesting enough.

Let  $\mathbb{T}$  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.<sup>27</sup> 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  $\mathcal{L}_{\Box K}$  contains names for each of its expressions, including terms, sentences and formulas (even those containing  $K$  and  $\Box$ ) and we also assume that we have some appropriate form of coding for the expressions of the language such that  $\ulcorner e \urcorner$  is the name of (the code of) the expression  $e$  (under this coding). Let  $\mathbb{T}_{\Box K}$  be the theory that results from adding the following axioms and rules to  $\mathbb{T}$ :

- |                   |   |  |
|-------------------|---|--|
| ( $\Box$ -fact)   | $\Box\varphi \rightarrow \varphi$   |  |
| ( $\Box$ -nec)    | If $\vdash_{\mathbb{T}_{\Box K}} \varphi$ , then $\vdash_{\mathbb{T}_{\Box K}} \Box\varphi$ |  |
| ( $\Box$ -dist)   | $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$             |  |
| ( $\Box$ -trans)  | $\Box\varphi \rightarrow \Box\Box\varphi$   |  |
| ( $\Box$ -Barcan) | $\forall x\Box\varphi(x) \rightarrow \Box\forall x\varphi(x)$                               |  |
| ( $\Box$ -ineq)   | $t_1 \neq t_2 \rightarrow \Box(t_1 \neq t_2)$   | for <i>rigid</i> terms $t_1$ and $t_2$ <sup>28</sup> |
| ( $K$ -fact)      | $K\varphi \rightarrow \varphi$  |  |
| ( $K$ -nec)       | If $\vdash_{\mathbb{T}_{\Box K}} \varphi$ , then $\vdash_{\mathbb{T}_{\Box K}} K\varphi$    |  |
| ( $K$ -dist)      | $K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$                      |  |
| ( $K$ -Barcan)    | $\forall xK\varphi(x) \rightarrow K\forall x\varphi(x)$                                     |  |
| ( $K$ -ineq)      | $t_1 \neq t_2 \rightarrow K(t_1 \neq t_2)$  | for <i>rigid</i> terms $t_1$ and $t_2$               |

<sup>27</sup> 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.

<sup>28</sup> We say that a term  $t$  is *rigid* if and only if the denotation of  $t$  is the same at every world.

In other words, the theory  $\mathbb{T}_{\square K}$  has an S4-notion of necessity and a T-notion of knowledge.<sup>29</sup> The theory can be interpreted as usual. Consider a Kripke frame  $\mathcal{F} = \langle W, R_{\square}, R_K \rangle$  such that  $W$  is a set of worlds,  $R_{\square}$  is a reflexive and transitive relation on  $W$  and  $R_K$  is a reflexive relation on  $W$ . To avoid unnecessary complications and, more specifically, to simplify the interpretations, we are going to make three stipulations. First, we assume that the domain  $D$  is the same in every world (this is why the Barcan Formula is one of the axioms above). 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.<sup>30</sup> Third, the language has enough names for the objects in  $D$ . So we can develop the semantics for the quantifiers substitutionally.<sup>31</sup>

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  $\mathbb{T}_{\square K}$ . The models for  $\mathbb{T}_{\square K}$  are of the form  $\langle W, R_{\square}, R_K, I^w \rangle$  where  $W, R_{\square}$  and  $R_K$  are as before and  $I^w$  is a function that assigns in each world  $w \in W$  an object  $c^{\mathcal{M}}$  in  $D$  to each individual constant  $c$ , a function  $f^{\mathcal{M}}$  from  $D^k$  to  $D$  to each  $k$ -ary function symbol  $f$ , and a relation  $r^{\mathcal{M}}$  in  $D^k$  to each  $k$ -ary relation symbol  $r$ . We can define an assignment of truth values to the formulas of  $\mathcal{L}_{\square K}$  relative to a model  $\mathcal{M}$  and a world  $w$  as follows:

- $(\mathcal{M}, w) \models r(t_1, \dots, t_k)$  if and only if  $\langle I^w(t_1), \dots, I^w(t_k) \rangle \in I^w(r)$
- $(\mathcal{M}, w) \models \neg \varphi$  if and only if  $(\mathcal{M}, w) \not\models \varphi$
- $(\mathcal{M}, w) \models \varphi_1 \wedge \varphi_2$  if and only if  $(\mathcal{M}, w) \models \varphi_1$  and  $(\mathcal{M}, w) \models \varphi_2$
- $(\mathcal{M}, w) \models \exists x \varphi$  if and only if  $(\mathcal{M}, w) \models \varphi_{[x/t]}$  for some term  $t$  of  $\mathcal{L}_{\square K}$
- $(\mathcal{M}, w) \models \square \varphi$  if and only if  $(\forall v \in W)(w R_{\square} v \rightarrow (\mathcal{M}, v) \models \varphi)$
- $(\mathcal{M}, w) \models K \varphi$  if and only if  $(\forall v \in W)(w R_K v \rightarrow (\mathcal{M}, v) \models \varphi)$

It can be shown that there are assignments of truth values that make all the axioms and rules for  $K$  and  $\square$  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.<sup>32</sup> So far, so good.

<sup>29</sup> There is nothing special about this particular choice of axioms, except that they keep the construction of models for the theory simple. In fact, some of the axioms seem quite controversial. For example, the axiom (K-ineq) requires that the agent be not only logically omniscient but also that she knows every true inequality. In any case, if the reader has other notions of necessity and knowledge in mind, a different set of axioms could be chosen.

<sup>30</sup> Some function symbols (such as successor, addition and multiplication) will of course denote the same function in every world.

<sup>31</sup> Nothing crucial depends on these simplifications. They just make the construction of models for the theory much easier.

<sup>32</sup> Actually things are not so simple. Since we have non-rigid terms in the language, some first order principles must be restricted. In particular,

$$(i) \varphi(t) \rightarrow \exists x \varphi(x)$$

The difficult part is to add a truth predicate (or a sequence  $T_1, T_2, T_3, \dots$  of truth predicates) to  $\mathcal{L}_{\Box K}$  and principles for truth to  $\mathbb{T}_{\Box K}$ . Let  $\mathcal{L}_{\Box K T_1(x)}$  be the language that results from adding the truth predicate  $T_1$  to  $\mathcal{L}_{\Box K}$ ,  $\mathcal{L}_{\Box K T_2(x)}$  the language that results from adding the truth predicate  $T_2$  to  $\mathcal{L}_{\Box K T_1(x)}$  and so on. Finally, let  $\mathcal{L}_{\Box K T_\omega(x)}$  be the language that results from adding the sequence of truth predicates  $T_1, T_2, T_3, \dots$  to  $\mathcal{L}_{\Box K}$ . Similarly, let  $\mathbb{T}_{\Box K T_\omega(x)}$  be the theory that results from adding the following axioms to  $\mathbb{T}_{\Box K}$  for each  $n \in \omega$  and for each predicate  $r$  of  $\mathcal{L}_{\Box K T_\omega(x)}$ :

$$\begin{aligned}
(\text{At}_{n+1}) \quad & \forall t_1, \dots, \forall t_k (\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}(\underline{r}(t_1, \dots, t_k)) \rightarrow (T_{n+1}(\underline{r}(t_1, \dots, t_k)) \\
& \leftrightarrow r(\text{val}^{\mathbb{W}}(t_1), \dots, \text{val}^{\mathbb{W}}(t_k))))^{33} \\
(\text{Comp}\neg_{n+1}) \quad & \forall y (\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}(y) \rightarrow (T_{n+1}(\neg y) \leftrightarrow \neg T_{n+1}(y))) \\
(\text{Comp}\wedge_{n+1}) \quad & \forall y \forall z (\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}(y \wedge z) \rightarrow (T_{n+1}(y \wedge z) \leftrightarrow T_{n+1}(y) \wedge T_{n+1}(z))) \\
(\text{Comp}\exists_{n+1}) \quad & \forall v \forall y (\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}(\exists v y) \rightarrow (T_{n+1}(\exists v y) \leftrightarrow \exists t (T_{n+1}(y(t/v))))^{34} \\
(\Box T_{n+1}) \quad & \forall y (\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}(y) \rightarrow (T_{n+1}(\Box y) \leftrightarrow \Box T_{n+1}(y))) \\
(K T_{n+1}) \quad & \forall y (\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}(y) \rightarrow (T_{n+1}(\underline{K}y) \leftrightarrow K T_{n+1}(y))) \\
(T T_{n+1}) \quad & \forall t (\text{Sent}_{\mathcal{L}_{\Box K T_{j-1}(x)}}(\text{val}^{\mathbb{W}}(t)) \rightarrow (T_{n+1}(\underline{T}_j(t)) \leftrightarrow T_{n+1}(\text{val}^{\mathbb{W}}(t)))) \\
& \text{for each } j \leq n^{35}
\end{aligned}$$

The predicate  $\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}(x)$  is true of exactly the (codes of the) sentences of  $\mathcal{L}_{\Box K T_n(x)}$ . So the truth axioms are restricted to the  $T_{n+1}$ -free part of the language (but they can contain occurrences of  $\Box$ ,  $K$  and  $T_j$  for  $j < n+1$ ). Of course, we need seven such axioms for each truth predicate in the language, so the theory  $\mathbb{T}_{\Box K T_\omega(x)}$  contains infinitely many  $(\text{At}_{n+1})$  axioms, infinitely many  $(\text{Comp}\neg_{n+1})$  axioms, and so on.

Also, once the truth predicates are part of the language, we can strengthen the schemas presented above for  $\Box$  and  $K$  by replacing them with their universally quanti-

$$(ii) s = t \rightarrow (\varphi(s) \leftrightarrow \varphi(t))$$

must restricted to rigid terms  $s$  and  $t$ . Otherwise it would be easy to come up with counterexamples. For the first, let  $\varphi(x)$  be “it is known that  $x$  is tall” and let  $t$  be “the president of Argentina”; and for the second let  $\varphi(x)$  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, 87ss).

<sup>33</sup> Here I want to be a little less sloppy with the notation. First,  $\text{val}^{\mathbb{W}}(x)$  is a function that, when applied to a closed term of  $\mathcal{L}_{\Box K T_n(x)}$ , gives the value of that term at world  $\mathbb{W}$ . We assume that the function  $\text{val}^{\mathbb{W}}(x)$  is primitive recursive, and so representable in  $\mathbb{T}_{\Box K T_\omega(x)}$ . Second, the expression  $\neg$  is not the negation sign but a function that applied to the code of a formula gives the code of its negation. Similarly, in  $T_{n+1}(\underline{r}(t_1, \dots, t_k))$ ,  $\underline{r}: \omega^k \rightarrow \omega$  is a function that applied to (the tuple of codes of)  $t_1, \dots, t_k$  outputs (the code of) the sentence  $r(t_1, \dots, t_k)$ . And likewise for the other expressions.

<sup>34</sup> Since we have stipulated that we have a term for every object, there is no need to quantify over objects or to use a satisfaction predicate in the right part of the biconditional.

<sup>35</sup> It would be redundant to add axioms of the form

$$(\text{TT}'_{n+1}) \quad \forall t (\text{Sent}_{\mathcal{L}_{\Box K T_{j-1}(x)}}(\text{val}^{\mathbb{W}}(t)) \rightarrow (T_{n+1}(\underline{T}_j(t)) \leftrightarrow T_j(\text{val}^{\mathbb{W}}(t))) \quad \text{for each } j \leq n.$$

These axioms are unnecessary, since they are cases of  $(\text{At}_n)$ .

fied versions. For example, instead of ( $\Box$ -fact) the theory will contain for each  $n$ , an axiom of the form  $\forall y(\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}(y) \rightarrow (\Box T_{n+1}(y) \rightarrow T_{n+1}(y)))$  and instead of ( $K$ -fact) the theory will contain for each  $n$ , an axiom of the form  $\forall y(\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}(y) \rightarrow (KT_{n+1}(y) \rightarrow T_{n+1}(y)))$ . So there is no need to use knowledge and necessity predicates to express quantifications involving those modalities. The complex predicates  $\Box T_n(x)$  and  $KT_n(x)$  (for each  $n$ ) do all the work. In this way, the theory is as expressive and as deductively powerful as the approach that treats modalities as predicates.<sup>36</sup>

In order to establish the consistency of the theory  $\mathbb{T}_{\Box K T_n(x)}$ , we now construct a natural model for it. The model is not only designed to show the consistency of the theory, but also to show that we can interpret the modal operators in a plausible way. It is unproblematic to assume that the theory  $\mathbb{T}_{\Box K}$  has models of the form  $\langle W, R_{\Box}, R_K, I^w \rangle$ . It is only the truth predicates that might generate a problem. A model  $\mathcal{M}$  for the theory  $\mathbb{T}_{\Box K T_n(x)}$  will be a structure of the form  $\langle W, R_{\Box}, R_K, I^w, \text{Ext} \rangle$ , where  $\text{Ext}: W \times \{T_n \mid n \in \omega\} \rightarrow \wp(\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}})$  is a function that assigns an extension to each truth predicate at each world  $w \in W$ . We impose the following condition on  $\text{Ext}$ :

For every  $w \in W$  and every  $\varphi$  in  $\mathcal{L}_{\Box K T_n(x)}$ ,  $(\mathcal{M}, w, \text{Ext}) \models T_{n+1} \ulcorner \varphi \urcorner$  if and only if  $(\mathcal{M}, w, \text{Ext}) \models \varphi$

Here the notation ' $(\mathcal{M}, w, \text{Ext}) \models \varphi$ ' is used to represent the fact that the formula  $\varphi$  is true relative to the model  $\mathcal{M}$ , the world  $w$  and the function  $\text{Ext}$ . It is clear that the structure  $\langle W, R_{\Box}, R_K, I^w, \text{Ext} \rangle$  will make every axiom of  $\mathbb{T}_{\Box K}$  true. It remains to show that the new axioms for truth are true in this structure (for each  $n$ ). We just present the proofs for the novel axioms ( $\Box T_{n+1}$ ), ( $KT_{n+1}$ ) and ( $\ulcorner T_{n+1} \urcorner$ ).

Proof for ( $\Box T_{n+1}$ ) and ( $KT_{n+1}$ ): Assume that for some model  $\mathcal{M}$  (with the features described), world  $w$  and function  $\text{Ext}$ ,  $(\mathcal{M}, w, \text{Ext}) \models T_{n+1} \ulcorner \Box \varphi \urcorner$  for some formula  $\varphi \in \text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}$ . By the definition of  $\text{Ext}$ , that holds if and only if  $(\mathcal{M}, w, \text{Ext}) \models \Box \varphi$ . Applying the semantic clause for  $\Box$ , we know that the previous fact holds if and only if  $(\mathcal{M}, v, \text{Ext}) \models \varphi$  for each world  $v$  such that  $wR_{\Box}v$ . This is the case if and only if  $(\mathcal{M}, v, \text{Ext}) \models T_{n+1} \ulcorner \varphi \urcorner$  for each world  $v$  such that  $wR_{\Box}v$ . In turn, by the semantic clause for  $\Box$ , this holds if and only if  $(\mathcal{M}, w, \text{Ext}) \models \Box T_{n+1} \ulcorner \varphi \urcorner$ .

For ( $KT_{n+1}$ ) just replace in the previous paragraph  $\Box$  everywhere by  $K$ .

Proof for ( $\ulcorner T_{n+1} \urcorner$ ): Assume that for some model  $\mathcal{M}$  (with the features described above), world  $w$  and function  $\text{Ext}$ ,  $(\mathcal{M}, w, \text{Ext}) \models T_{n+1} \ulcorner \ulcorner T_i \urcorner(t) \urcorner$  for some term  $t$  such that  $\text{val}^w(t) = \varphi$  for some formula  $\varphi \in \text{Sent}_{\mathcal{L}_{\Box K T_{j-1}(x)}}$ . By the definition of  $\text{Ext}$ , this holds if and only if  $(\mathcal{M}, w, \text{Ext}) \models T_j \ulcorner \varphi \urcorner$ . And this is the case if and only if  $(\mathcal{M}, w, \text{Ext}) \models \varphi$ , which holds if and only if  $(\mathcal{M}, w, \text{Ext}) \models T_{n+1} \ulcorner \varphi \urcorner$ .

---

<sup>36</sup> A technical argument supporting this claim is developed in (Halbach and Welch 2009).

This is enough to establish:

(Theorem 5)  $\mathbb{T}_{\Box K T_\omega(x)}$  is consistent.

It can be shown that  $\mathbb{T}_{\Box K T_\omega(x)}$  has other attractive features in addition to being consistent. For example, we can recover the T-schema for well behaved sentences:

(Theorem 6) If a formula  $\varphi$  has no occurrences of  $T_j$  for  $j > n$ , then  
 $\vdash_{\mathbb{T}_{\Box K T_\omega(x)}} T_{n+1} \ulcorner \varphi \urcorner \leftrightarrow \varphi$  for each  $n \in \omega$ .

Proof: This can be proved by a simple induction on the complexity of  $\varphi$ , as the reader can check for herself.

It is clear from the previous theorem that the T-schema does not hold unrestrictedly. For instance,  $T_n \ulcorner \varphi \urcorner \leftrightarrow \varphi$  is not a theorem of  $\mathbb{T}_{\Box K T_\omega(x)}$  if  $\varphi$  contains an occurrence of  $T_n$ . This shows that multimodal paradoxes are no longer a problem. Although a sentence such as  $A \leftrightarrow \neg \Box T_n \ulcorner K T_n \ulcorner A \urcorner \urcorner$  is provable in the theory using the Diagonal Lemma, no contradiction follows from it, since  $A$  involves an occurrence of  $T_n$ .

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:

(Theorem 7) If a formula  $\varphi$  has no occurrences of  $T_j$  for  $j > n$ , then:

$$\begin{aligned} &\vdash_{\mathbb{T}_{\Box K T_\omega(x)}} \Box T_{n+1} \ulcorner \varphi \urcorner \leftrightarrow \Box \varphi, \text{ and} \\ &\vdash_{\mathbb{T}_{\Box K T_\omega(x)}} K T_{n+1} \ulcorner \varphi \urcorner \leftrightarrow K \varphi.^{37} \end{aligned}$$

### 8. Capturing the Knowability Argument

The theory  $\mathbb{T}_{\Box K T_\omega(x)}$  has two interesting features that might be relevant to properly evaluate the Knowability Argument. First, it could be argued that using schematic versions of the relevant principles in the argument is unsatisfactory, especially when these principles are available in quantificational form. So the reasoning should go through even for quantified versions of the principles. Second, the Knowability Principle says that every *truth* is knowable. So it should be formalized using the truth predicate. Taking into consideration these two observations, we formalize the Knowability Principle in the following way:

$$(KP_{n+1}) \quad \forall y (\text{Sent}_{\Box K T_n(x)}(y) \rightarrow (T_{n+1}(y) \rightarrow \Diamond K T_{n+1}(y)))^{38}$$

And actually, there is one statement of this form for each  $n$ .

<sup>37</sup> This last theorem might seem controversial since some philosophers argue that an agent can know something without knowing it is true. However, pursuing this issue here would take me too far from the point I am trying to make. A careful discussion is provided in (Pailos & Rosenblatt 2013).

<sup>38</sup> The reason for adding a truth predicate in the consequent as well is that in our theory  $K T(x)$  plays the role of the knowledge predicate (actually, we should formalize the consequent as  $\Diamond T(\underline{K T}(x))$ , but we will omit the second truth predicate to simplify things).

Is it possible to reconstruct Fitch's derivation using this principle and other principles available in our theory? The answer is yes. This is immediate from the fact that for every appropriate formula  $\varphi$  and each  $n$  we can replace  $\top_{n+1} \ulcorner \varphi \urcorner$  and  $\varphi$  in every context. A less direct way of checking that the derivation goes through is the following.<sup>39</sup>

Assume that for some model  $\mathcal{M}$  of the theory  $\mathbb{T}_{\Box K T_n(x)}$ , world  $\mathfrak{w}$  and function  $\text{Ext}$  it is the case that  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \forall y (\text{Sent}_{\mathcal{L}_{\Box K T_n(x)}}(y) \rightarrow (\top_{n+1}(y) \rightarrow \Diamond K \top_{n+1}(y)))$  for some  $n > 1$ . Assuming that  $\top_n \ulcorner \varphi \urcorner \wedge \neg K \top_n \ulcorner \varphi \urcorner$  is a sentence of  $\mathcal{L}_{\Box K T_n(x)}$  for some sentence  $\varphi$  of  $\mathcal{L}_{\Box K T_{n-1}(x)}$ , this implies that:

$$(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \top_{n+1} \ulcorner \top_n \ulcorner \varphi \urcorner \wedge \neg K \top_n \ulcorner \varphi \urcorner \urcorner \rightarrow \Diamond K \top_{n+1} \ulcorner \top_n \ulcorner \varphi \urcorner \wedge \neg K \top_n \ulcorner \varphi \urcorner \urcorner.$$

However, it can be shown that  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \neg \Diamond K \top_{n+1} \ulcorner \top_n \ulcorner \varphi \urcorner \wedge \neg K \top_n \ulcorner \varphi \urcorner \urcorner$ . For suppose that  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models K \top_{n+1} \ulcorner \top_n \ulcorner \varphi \urcorner \wedge \neg K \top_n \ulcorner \varphi \urcorner \urcorner$ . Then  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \top_{n+1} \ulcorner \top_n \ulcorner \varphi \urcorner \wedge \neg K \top_n \ulcorner \varphi \urcorner \urcorner$  for every  $\mathfrak{v}$  such that  $\mathfrak{w} R_K \mathfrak{v}$ . By the definition of  $\text{Ext}$ ,  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \top_n \ulcorner \varphi \urcorner \wedge \neg K \top_n \ulcorner \varphi \urcorner$  for every  $\mathfrak{v}$  such that  $\mathfrak{w} R_K \mathfrak{v}$ , so  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \top_n \ulcorner \varphi \urcorner$  for every  $\mathfrak{v}$  such that  $\mathfrak{w} R_K \mathfrak{v}$ , and  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \neg K \top_n \ulcorner \varphi \urcorner$  for every  $\mathfrak{v}$  such that  $\mathfrak{w} R_K \mathfrak{v}$ . This means that  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models K \neg K \top_n \ulcorner \varphi \urcorner$  and  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models K \neg K \top_n \ulcorner \varphi \urcorner$ . By the reflexivity of  $R_K$ , this last claim implies that  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \neg K \top_n \ulcorner \varphi \urcorner$ . And this is a contradiction. So, as we said before,  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \neg \Diamond K \top_{n+1} \ulcorner \top_n \ulcorner \varphi \urcorner \wedge \neg K \top_n \ulcorner \varphi \urcorner \urcorner$ . But then we can infer that  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \neg \top_{n+1} \ulcorner \top_n \ulcorner \varphi \urcorner \wedge \neg K \top_n \ulcorner \varphi \urcorner \urcorner$ . By  $\text{Ext}$ , this implies that  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \neg (\top_n \ulcorner \varphi \urcorner \wedge \neg K \top_n \ulcorner \varphi \urcorner)$ . It follows that  $(\mathcal{M}, \mathfrak{w}, \text{Ext}) \models \top_n \ulcorner \varphi \urcorner \rightarrow K \top_n \ulcorner \varphi \urcorner$ . Since this holds for any  $n > 0$ , we can infer that each  $(\text{KP}_n)$  is false.

### 9. Closing remarks

The reasoning in the last section shows that  $\mathbb{T}_{\Box K T_n(x)}$  is capable of representing Fitch's reasoning and that adding the Knowability Principle to the theory  $\mathbb{T}_{\Box K T_n(x)}$  leads to something clearly false. So the best theory the type-theorist has available for solving other well-known paradoxes does not block the Knowability Argument. This is an advantage, I think, over theories that type truth, knowledge and other modal predicates. Although the typing of knowledge is useful to block both Fitch's reasoning and Montague's paradox for knowledge, the truth-types are still needed to avoid truth-theoretic paradoxes. Since knowledge and truth are both typed, either we run into multimodal paradoxes or we have to set up some unattractive mixed hierarchy of predicates. Neither alternative seem satisfactory. Hence, it seems that if the Knowability Argument is to be rejected, the best way to do that is not by means of knowledge-types.

<sup>39</sup> For a similar reasoning see (Horsten 2009).

Moreover, the demand that paradoxes of the same family should receive the same sort of solution is satisfied only to some extent on the approach that has types for every predicate. While it is true the every paradox is avoided by the use of types, truth-theoretic paradoxes require truth-types, the paradoxes of knowledge require knowledge-types, and so on. On the account that emulates modal predicates by using truth, every paradox is avoided by the use of truth-types.

## REFERENCES

- Carrara, Massimiliano and Davide Fassio. 2011. Why Knowledge Should Not Be Typed: An Argument against the Type Solution to the Knowability Paradox. *Theoria* 77: 180-193.
- Fagin, Ronald, Joseph, Halpern, Yoram, Moses, and Moshe Vardi. 1995. *Reasoning about Knowledge*. Cambridge, MA: MIT Press.
- Fitch, Frederic. 1963. A logical analysis of some value concepts. *The Journal of Symbolic Logic* 28: 135-42.
- Halbach, Volker. 2011. *Axiomatic theories of truth*. New York: Cambridge University Press.
- . 2008. On a side effect of solving Fitch's paradox by typing knowledge. *Analysis* 68: 114-120.
- and Philip Welch. 2009. Necessities and Necessary Truths: A Prolegomenon to the Use of Modal Logic in the Analysis of Intensional Notions. *Mind* 118: 71-100.
- Hart, William. 2009. Invincible Ignorance. In *New essays on the knowability paradox*, ed. Joe Salerno, 320-323. New York: Oxford University Press.
- Horsten, Leon. 2009. Levity. *Mind* 118: 555-581.
- Jenkins, Carrie. 2009. The Mystery of the Disappearing Diamond. In *New essays on the knowability paradox*, ed. Joe Salerno, 302-319. New York: Oxford University Press.
- Kripke, Saul. 1975. Outline of a theory of truth. *The Journal of Philosophy* 72: 690-716.
- Kvanvig, Jonathan. 2008. *The Knowability Paradox*. Oxford: Oxford University Press.
- Linsky, Bernard. 2009. Logical types in some arguments about knowability and belief. In *New essays on the knowability paradox*, ed. Joe Salerno, 163-181. New York: Oxford University Press.
- Montague, Richard. 1963. Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica* 16: 155-67. (Reprinted in Montague 1974).
- . 1974. *Formal Philosophy: Selected Papers of Richard Montague*. New Haven and London: Yale University Press.
- Pailos, Federico and Lucas Rosenblatt. Forthcoming 2014. Solving Multimodal Paradoxes. *Theoria*.
- Paseau, Alexander. 2009. How to type: reply to Halbach. *Analysis* 69: 280-286.
- . 2008. Fitch's argument and typing knowledge. *Notre Dame Journal of Formal Logic* 49: 155-176.
- Salerno, Joe, ed. 2009. *New essays on the knowability paradox*, New York: Oxford University Press.
- Priest, Graham. 2009. Beyond the Limits of Knowledge. In *New essays on the knowability paradox*, ed. Joe Salerno, 93-104. New York: Oxford University Press.

**LUCAS ROSENBLATT** is Teaching Assistant and PhD Student at the University of Buenos Aires. He writes mainly on issues in philosophical logic, especially on paradoxes and non-classical logic.

**ADDRESS:** Puan 480, Facultad de Filosofía y Letras, University of Buenos Aires, Ciudad de Buenos Aires, Argentina. E-mail: l\_rosenblatt@hotmail.com