Benedikt **Löwe**, Wolfgang **Malzkorn**, Thoralf **Räsch** (*eds.*)
Foundations of the Formal Sciences II
Applications of Mathematical Logic in Philosophy and Linguistics
Papers of the conference held in Bonn, November 10-13, 2000

The Logic of Intensional Predicates

Leon Horsten*
University of Leuven
Department of Philosophy
Kardinaal Mercierplein 2
3000 Leuven
Belgium
E-mail: `Leon.Horsten@hiw.kuleuven.ac.be`

Abstract. This paper gives a personal overview of the current situation concerning the logical theory of intensional predicates. It is shown how several intensional notions, when logically treated as predicates, yield liar-like paradoxes. Some consistent semantic and axiomatic theories of intensional predicates are presented and discussed. To conclude, an inquiry is made into the conceptual relation between truth and intensional notions.

1 Introduction

In philosophical logic, it is standard to formalise intensional notions such as necessity, knowledge, past / future as propositional operators instead of as predicates. Nevertheless, the prima facie desirability of logically treating intensional notions as predicates is well-known. We routinely say things like 'There is something of importance that Carl does not know', *i.e.*, we quantify over the things (whatever they may be) of which necessity, knowledge etc. is *predicated*. If we formalise intensional notions as

Received: March 30th, 2001;
Accepted by the editors: June 7th, 2001.
2000 *Mathematics Subject Classification.* **03B45 03A05** 03B10 03B60.

* I am grateful to Benedikt Löwe for inviting me to present a paper at the Foundations of the Formal Sciences II conference. Thanks to Volker Halbach, Hannes Leitgeb and Philip Welch for helpful comments and for information about their joint work in progress.

sentential operators, then there appears to be no natural way to achieve such quantification in the formal language.[1]

Unfortunately, several intensional notions, when formally treated as predicates, give rise to liar-like paradoxes. In this sense, the logical behavior of these predicates is similar to that of the concept of truth. The propositional operator approach, in contrast, yields consistent systems, for which interesting semantics has been developed (Kripke semantics). For this reason, logicians have opted for formalising intensional notions as propositional operators. If the corresponding operator for truth would not have been trivial but would have given rise to interesting formal systems, then undoubtedly many logicians would have followed the same policy with respect to the notion of truth.

The logical behavior of the truth predicate has been investigated intensively since the 1930s, both from a proof-theoretic and from a semantic point of view. Until recently, almost all of the predicate treatments of intensional notions that were proposed could be seen as fairly straightforward transpositions or adaptations of logical approaches to the notion of truth. Recently, this situation has started to change. The investigation of the logic of intensional predicates appears to be a nascent research domain. But, as we will see, it still remains a challenge to develop really new and appropriate tools for the logical treatment of intensional predicates.

This paper is structured as follows. Section 2 contains an overview of the known *negative* results about possible treatments of intensional notions as predicates. In Section 3, the fruitful interplay that has taken place over the past decades between semantical and axiomatic approaches to the logic of truth is briefly revisited. Subsequently, attempts to *consistently* treat intensional notions as predicates are discussed. In Section 4, two semantical treatments of intensional predicates are reviewed. First, a Kripkean approach to modelling the logical interaction between knowability and truth, using iterated inductive definitions, is sketched. Secondly, a sketch is given of a recently developed possible worlds semantics for formal languages containing a necessity predicate. Section 5 looks at proof-theoretic approaches to the theory of knowability treated as a predicate. The concluding section takes up the philosophical question whether

[1] On the operator approach, the desired forms of quantification could be realised by introducing *propositional quantifiers* into the formal language. But then, given sufficient expressive power, liar-like paradoxes can be generated, as in the predicate approach. For details, see [Gri93].

an explanation can be given of the striking formal similarity between the paradoxicality of the truth predicate and the paradoxicality of certain intensional notions.

This paper concentrates on guiding ideas rather than on technical details. Nevertheless, numerous references to sources where these details are spelled out are supplied, often in footnotes. A conscious attempt has been made to mention open problems, and to hint at directions which future logical research on intensional predicates could take. This paper does not aim at giving anything like a complete overview of predicate approaches to intensional notions. Also, no claim is made concerning the relative importance of the approaches discussed in this paper. The selection is merely a reflection of my personal taste and my knowledge of the field.

The background knowledge required for reading this paper is quite modest. The reader is assumed to be familiar with what is involved in elementary Gödelian incompleteness phenomena (especially the diagonal lemma), with elementary facts about Kripke semantics for (operator) modal logic, and is assumed to have a rudimentary knowledge of the main logico-philosophical theories of truth and the semantic paradoxes. The use of notation in this paper should be sufficiently standard. In the interest of readability, a certain looseness in notation is at some places exhibited (especially in Section 5). It is assumed that the reader can supply the required precision if he so desires.

2 Intensional Paradoxes

Around 1960, it became clear that the intensional notion of necessity and the notion of knowability by a fixed agent, when treated as predicates, give rise to liar-like paradoxes. These paradoxes are now known as the *Paradox of the Knower*[2] and the *Paradox of Necessity*.[3] Formally, the arguments for these two paradoxes have the same structure. Take the language of arithmetic, augmented with a new primitive predicate P. Call this language $\mathcal{L}_{\mathrm{P}}$. For definiteness, let us interpret P as "it is knowable that". Then it can be shown that the theory T_1 consisting of (the first-order closure of):

[2] Due to Kaplan and Montague, 1960. The paradox was also, seemingly independently, discovered by Myhill. *Cf.* [Myh60].

[3] *Cf.* [Mon63].

- the (finitely axiomatized) weak arithmetical theory Q (formulated in $\mathcal{L}_{\mathrm{P}}$),
- $\mathrm{P}\overline{A} \rightarrow A$,[4]
- $A \Rightarrow \mathrm{P}\overline{A}$.

is inconsistent. The epistemic axiom scheme is called *Reflexivity*. The epistemic rule of inference is called the rule of *Necessitation* (or may be called *Provabilitation*, if the epistemic reading is intended).

The argument showing T_1 to be inconsistent is exceedingly simple:

Proof. Gödel's diagonal lemma is provable in Q for the extended language $\mathcal{L}_{\mathrm{P}}$. Therefore there exists a sentence G such that $T_1 \vdash G \leftrightarrow \neg\mathrm{P}\overline{G}$. We now reason in T_1. Suppose $\neg G$. Then, by the instance of the diagonal lemma, $\mathrm{P}\overline{G}$. Therefore, by the Reflexivity axiom, G follows. Contradiction. So $T_1 \vdash G$. By the Necessitation rule, $T_1 \vdash \mathrm{P}\overline{G}$. So by the instance of the diagonal lemma again, $T_1 \vdash \neg G$. Contradiction.

Yet this conclusion is genuinely paradoxical: all of the axioms of T_1 intuitively appear sound on the intended interpretation of P as either necessity or as knowability.

Thomason later argued that the intensional notion "it is rational (for a fixed agent) to believe that" is for similar reasons paradoxical.[5] Specifically, he shows, using a diagonal argument, that the theory T_2 consisting of:

- Q
- $\mathrm{P}(\overline{\mathrm{Q}})$
- $\mathrm{P}\overline{\mathrm{P}\overline{A} \rightarrow A}$
- $A \Rightarrow \mathrm{P}\overline{A}$
- $\mathrm{P}\overline{A} \rightarrow \mathrm{P}\overline{\mathrm{P}\overline{A}}$
- $\mathrm{P}\overline{A} \rightarrow (\mathrm{P}\overline{A \rightarrow B} \rightarrow \mathrm{P}\overline{B})$

is inconsistent. The claim here again is that on the interpretation of P as "it is rational to believe that", T_2 intuitively appears to be sound for its intended interpretation. However, there appears to be some doubt whether Thomason has provided us with a genuine paradox. Specifically, one may

[4] $\overline{A}$ stands for the Gödel number of A.

[5] See [Thoo80].

wonder whether the axiom $\mathrm{P}\overline{\mathrm{P}\overline{A} \to A}$ is intuitively plausible. We sometimes find ourselves in a situation where we have good reasons to believe A, but where A is nevertheless false. We know that this is so. Would it not therefore be more rational to suspend judgement on many sentences of the form $\mathrm{P}\overline{A} \to A$ than to believe in all of them?

Recently it became clear that the temporal predicates “it has always been the case that” and “it will always be the case that” are also subject to paradoxes.[6] Consider the language of arithmetic augmented with two new primitive predicates, H (“it has always been the case that”) and G (“it will always be the case that”). As in ordinary tense logic, the notions “at some moment in the past, it was the case that” (P) and “at some moment in the future, it will be the case that” (F) can be defined in the usual way in terms of these predicates H and G and negation. It can then be shown using a diagonal argument that, in the context of the theory Q, the predicate version T_3 of Prior’s minimal tense logic K_t is *internally inconsistent*, *i.e.*, it proves $\mathrm{H}\overline{\bot}$ and $\mathrm{G}\overline{\bot}$ (where $\bot$ is the falsum symbol). Clearly this is counterintuitive. Moreover, it can be shown, again on the basis of a Gödelian diagonal argument, that when the axiom $\mathrm{H}\overline{A} \to \mathrm{P}\overline{A}$ (which will be called HP) is added to a fragment of T_3, an outright inconsistency is reached. This axiom says, roughly, that time has a a first moment.

The axiom HP has some *empirical* support: on some interpretations of the Big Bang model of the General Theory of Relativity, it is true. At the same time, it is clear that HP falls short of qualifying as a *conceptual* truth about the notion of past and future. Nevertheless, it seems that it is not the business of (tense) *logic* to exclude the possibility that there is a first moment in time. So the inconsistency that has been reached is paradoxical.

Since the middle 1980s, there exists also a class of theorems which show that certain apparently weak axiomatic theories of truth are nevertheless in some sense paradoxical, *i.e.*, inconsistent, internally inconsistent or ω-inconsistent. In some cases these results also generate paradoxes for intensional notions. For instance, consider the pure past fragment T_3^- of the theory T_3, *i.e.*, the fragment of T_3 consisting of the formulas in which the predicate G does not occur. If the axiom HP is added to T_3^-, a consistent system T_4 is obtained, since this system is a subsys-

[6] *Cf.* [Hor$_1$Lei01].

tem of the consistent system F of Friedman and Sheard. But Friedman and Sheard showed that if to T_4 the axiom

$$\mathrm{H}\overline{A} \rightarrow \mathrm{H}\overline{\mathrm{H}\overline{A}}$$

is added, a contradiction follows.[7] Also, a result of McGee[8] entails that if the Barcan formula is added to T_4, an inconsistency follows. Friedman and Sheard and McGee were in the first place interested in formal theories of truth. But they explicitly recognise the possibility that their results might also be relevant for other interpretations of the nonarithmetical primitive predicate.

This situation raises some questions. First, how far does this phenomenon extend? Might there be more intensional notions which are subject to similar paradoxes? For instance, one might want to look at the deontic notions. Second, one may wonder whether a deeper explication can be given of why these intensional notions and the apparently nonintensional notion of truth are subject to similar paradoxes. We will return to this second question towards the end of the paper.

3 Semantic and axiomatic theories of truth

The results of the previous section illustrate that over the years a considerable amount of information about the limits to possible theories concerning truth and intensional predicates has been collected. Let us now turn to positive results concerning possible theories of truth and intensional notions treated as predicates. First, we briefly recall some major positive theories of truth that have over the years been proposed.

In the empirical sciences, when one wants to develop a theory concerning a field of inquiry, one can either try to explicate basic laws concerning the phenomena under investigation, or one can try to construct models in which salient features of the behavior of these phenomena are represented. So it is in logic. In the theory of truth, there has been a fruitful interplay between semantical and axiomatic approaches.

The earliest theories of truth worthy of the name were Tarski's famous theories of truth: the disquotational theory and the compositional

[7] *Cf.* [FriShe87, p. 14].
[8] *Cf.* [McG85].

theory.[9] The distinguishing feature of these theories was that iterations of the truth predicate were banned, since these were deemed responsible for the paradoxes. Tarski's theories can be viewed as axiomatic theories of truth —although Tarski himself would not have wanted to put it this way. Using the notion of truth in a model, also discovered by Tarski, consistency proofs of these axiomatic theories of truth can be obtained.

Four decades later, Kripke constructed partial models for languages in which iterations of the truth predicate are admitted.[10] Two of Kripke's constructions were the inductive fixed point construction based on the strong Kleene scheme for evaluating formulas and the inductive fixed point construction based on the supervaluation scheme for evaluating formulas. These two semantic theories naturally lead to the formulation of two important axiomatic theories of truth. On the one hand, Feferman's system KF was directly motivated by the Kleene construction.[11] On the other hand, Cantini constructed a natural formalization, called VF, of the supervaluation approach.[12]

There exist cases where, even though the relevant connections were not initially seen, it later became clear that there are deep relations between certain axiomatic theories of truth and certain semantical approaches. One is reminded here of Halbach's observation of the relation between the axiomatic theory FS of Friedman and Sheard and the revision semantics of Gupta, Belnap and Herzberger: models of FS can in a sense be seen as natural revision models.[13]

What has been accomplished concerning theories of intensional notions is much less impressive. As mentioned before, the challenge in this area is to develop natural constructions which are not obvious adaptations of constructions for truth.

4 Semantic theories of intensional predicates

4.1 Kripkean approaches to the semantics of knowability

Mostly, what has been done for intensional predicates is to borrow ideas from theories of truth and to apply them to the investigation of the logic

[9] *Cf.* [Tar56a].
[10] *Cf.* [$\text{Kri}_1$75].
[11] *Cf.* [Fef91].
[12] *Cf.* [Can90].
[13] *Cf.* [$\text{Hal}_0$94].

of intensional notions. For instance, one can in a straightforward way model necessity and knowability as partial predicates in a Kripkean vein.[14] But this yields nothing really new. If anything, the resulting models for necessity and knowability are considerably less natural than their counterparts for truth.

An attempt to go beyond this consists in investigating the logical interaction between truth and knowability in a Kripkean framework. One tries to formulate Kripkean inductive rules for truth as a partial predicate and knowability as a partial predicate which exhibit a form of logical interaction between truth and knowability, or, more specifically, between truth and unknowability. Let us look at these rules in some more detail.[15]

We consider an arithmetical language with two partial predicates, T (truth) and B (knowability). We will construct two iterated hierarchy rules for T and B. First, we construct an auxiliary hierarchy rule $\mathcal{R}_1$. With the aid of $\mathcal{R}_1$, we then construct our main hierarchy rules $\mathcal{R}_2$ and $\mathcal{R}_3$.

The auxiliary hierarchy rule $\mathcal{R}_1$ is constructed as follows. We start by formulating natural Kripkean inductive clauses for T. We opt for the supervaluation approach. The next question is how the partial interpretation of B should be constructed. For the extension of B, we are immediately faced with a philosophical problem. It is hard to see what it would mean for the extension of B to grow even in the first transfinite stage ω. The extension of B grows in time, and this clearly is not the 'space' or 'dimension' in which the extension of T grows. So it seems that the Kripkean inductive framework does not allow us to express interesting features of the 'evolution' of the extension of B. Therefore, let the extension of B be constant in all stages. Set it equal to *Peano Arithmetic*, formulated in the extended language (which we abbreviate as PA), for instance. The anti-extension of B turns out to be more interesting. Everything that is false is definitely unknowable. Therefore the anti-extension of T must be a subset of the anti-extension of B. In the auxiliary rule $\mathcal{R}_1$ that we are constructing, set the anti-extension of B equal to the anti-extension of T. This completes the definition of $\mathcal{R}_1$. $\mathcal{R}_1$ intuitively appears sound and secure. And it can be formally verified that $\mathcal{R}_1$ is logically well-behaved. It is monotone, has a least fixed point, etc. It can be shown that in the

---

[14] *Cf., e.g.*, [Mor$_1$86].

[15] The approach sketched here is worked out in more detail in [Hor$_1$98].

least fixed point the liar sentence L, the absolute Gödel sentence G and the knower sentence K, are all indeterminate with respect to truth and knowability: they are neither determinately true or false, nor determinately knowable or unknowable. In sum, $\mathcal{R}_1$ is an ostensibly sound but not very exciting inductive rule.

Now we construct our first main hierarchy rule $\mathcal{R}_2$. This inductive hierarchy rule is just like $\mathcal{R}_1$, except for the successor clause for the anti-extension of B. At each successor stage $\alpha + 1$, we of course put in the anti-extension of B everything that is in the anti-extension of T at stage $\alpha + 1$. But aside from that, we put more sentences in the anti-extension of B. We consider each sentence φ, and ask the following question:

- **If** φ would be added to the extension of T at stage α, and from stage $\alpha+1$ onward the hierarchy would be continued according to the hierarchy rules of $\mathcal{R}_1$, **would** then at some later stage β an inconsistency arise in the extension of T?

If the answer to this question is yes, then φ is put in the anti-extension of B at stage $\alpha + 1$. If no, we go on to consider the next sentence. In this way, $\mathcal{R}_1$ is used as an auxiliary rule to co-determine the anti-extension of B at successor stages of $\mathcal{R}_2$.

The motivation behind this new successor clause is simply this: *if it is inconsistent, according to some unobjectionable rule, to consider a sentence true, then it is definitely unknowable*. Or, shorter still: if a sentence cannot be true, then it cannot be knowable. This may sound plausible enough, but experience with the paradoxes has taught us to be extremely apprehensive about such arguments. After all, it is inconsistent to consider the liar sentence L to be true. But if, on the strength of this, we would put L in the anti-extension of T, an inconsistency would arise! We need assurance that something similar does not happen for $\mathcal{R}_2$. Fortunately, it can be shown that $\mathcal{R}_2$ is logically well-behaved: it is monotone, consistent, has a least fixed point, etc.

Define the *liar sentence* as the sentence L such that PA $\vdash L \leftrightarrow \neg \mathrm{T}\overline{L}$, the *absolute Gödel sentence* G as the sentence such that PA $\vdash G \leftrightarrow \neg \mathrm{B}\overline{G}$, and the *knower sentence* as the sentence K such that PA $\vdash K \leftrightarrow \mathrm{B}\overline{\neg K}$. Since the diagonal lemma holds for PA, such sentences exist. It can be shown that the least fixed point of $\mathcal{R}_2$ classifies L, $\neg G$, and K as determinately unknowable. L, G, and K still do not receive a truth-value at the least fixed point of $\mathcal{R}_2$. For L this is as it should be. One may wonder whether G should not really be in the extension of T, and K in the anti-extension of B. Could there not be 'absolute' counterparts of Gödel's arguments[16] for the truth of the Gödel sentence for PA and the falsehood of the so-called Jeroslow sentence for PA which validate G, and falsify K?

Let us suspend this question for a moment and move on to the construction of our second main hierarchy rule $\mathcal{R}_3$. This rule also uses $\mathcal{R}_1$ as an auxiliary rule. $\mathcal{R}_3$ is just like $\mathcal{R}_2$, except for one small change in the successor clause for the anti-extension of B. At successor stages $\alpha + 1$ we again include the anti-extension of T at $\alpha + 1$, and we consider each sentence φ in the light of a question:

- If φ would be added to the extension of B at stage α, and from stage $\alpha + 1$ onward the hierarchy would be continued according to the hierarchy rules of $\mathcal{R}_1$, would then at some later stage β an inconsistency arise in the extension of T?

Again, if the answer is yes, then we put φ in the anti-extension of B at stage $\alpha + 1$, otherwise we move on to the next sentence.

It is easily seen that $\mathcal{R}_3$ is at least as strong as $\mathcal{R}_2$. $\mathcal{R}_3$ is motivated by saying that *if it is inconsistent to assume that a given sentence φ is knowable, then it is definitely unknowable.* This motivation is philosophically somewhat weaker than the motivation that was given for $\mathcal{R}_2$. Still, it seems that if we can coherently get away with this requirement, this would be a good thing. And fortunately, one can again verify that $\mathcal{R}_3$ is logically well-behaved. The interesting thing is that in the least fixed point of $\mathcal{R}_3$, G is in the extension of the truth predicate, thus validating an 'absolute' version of Gödel's argument for the truth of the Gödel sentence for PA, and K is in the anti-extension of the truth predicate.

In favor of the above Kripkean construction, one might say that it at least involves a new element. It essentially makes use of the theory of *iterated* inductive definitions, which is not usually done in Kripkean theories of truth. Nevertheless, the above construction has one markedly

[16] Gödel's argument for the truth of the Gödel sentence G_{PA} for PA (for which the diagonal property $\mathsf{PA} \vdash G_{\mathsf{PA}} \leftrightarrow \neg \mathrm{Bew}_{\mathsf{PA}}(\overline{G_{\mathsf{PA}}})$ holds) goes roughly as follows. Suppose G_{PA} is false. Then its negation is true, so that by the diagonal property G_{PA} is provable in PA. But whatever is provable in PA is true. Therefore G_{PA}, which contradicts our hypothesis. So G_{PA} must be true. By a similar line of reasoning, it can be shown that the Jeroslow sentence for PA must be false.

unattractive aspect, *i.e.*, the *asymmetric* treatment of the extension and the anti-extension of the knowability predicate.

4.2 Possible worlds semantics for intensional predicates

We do have, of course, since the work of Kripke in the early 1960s the successful program of constructing possible worlds models for intensional notions treated as sentential operators.[17] But in this approach, the problems of the paradoxes for intensional notions are simply brushed under the carpet: the language simply does not have enough expressive power to form the relevant diagonal sentences. And until recently, no one saw how results and techniques from this area could be used to construct interesting models for intensional predicates.

Now Halbach, Leitgeb and Welch have managed to construct an interesting theory of possible worlds models for necessity treated as a predicate. We here give the flavor of their approach.[18]

Consider the language of arithmetic, augmented with a new primitive necessity predicate $\Box$. In the familiar way, a possibility predicate $\Diamond$ can be defined in terms of negation and $\Box$. Call this language $\mathcal{L}_\Box$. At each possible world, the arithmetical portion of $\mathcal{L}_\Box$ is interpreted by the standard model $\mathbb{N}$ of the natural numbers. Only the set of (codes of) sentences which functions as extension of the necessity predicate is allowed to differ from possible world to possible world. A formula A of $\mathcal{L}_\Box$ is then said to be *true (false) at a possible world* if and only if A is true (false) in the classical model $\langle \mathbb{N}, X \rangle$ associated with this possible world, where X is the extension of $\Box$. We can abbreviate this as $X \models A$ ($X \not\models A$). A *frame* is defined (as in ordinary operator modal logic) as an ordered pair $\langle W, R \rangle$, where $W \neq \emptyset$ is a set of possible worlds, and R, the accessibility relation, is a binary relation on W. A *possible worlds model* $\mathcal{M}$ is then defined as a triple $\langle W, R, V \rangle$, such that $\langle W, R \rangle$ is a frame, and V is an assignment function which assigns to each $w \in W$ a subset of $\mathcal{L}_\Box$ such that

$$V(w) = \{A \in \mathcal{L}_\Box \mid \forall u\, (wRu \Rightarrow V(u) \models A)\}.$$

In words: the necessary truths at a world w consist of those sentences which are true at all worlds accessible from w.

[17] A good introduction to possible worlds semantics for modal logic is [Hug$_1$Cre96].

[18] For details, consult [Hal$_0$LeiWel∞].

It can then easily be shown, as in ordinary operator modal logic, that all possible worlds models verify the *normality* principles of Necessitation and Distribution:

Proposition 1. Suppose $\langle W, R, V \rangle$ is a possible worlds model, $w \in W$ and $A, B \in \mathcal{L}_{\Box}$. Then the following holds:
a. If $V(u) \models A$ for all $u \in W$, then $V(w) \models \Box A$.
b. $V(w) \models \Box\overline{A \rightarrow B} \rightarrow (\Box\overline{A} \rightarrow \Box\overline{B})$.

Likewise, as in operator modal logic, transitive frames validate the S4 axiom:

Proposition 2. Let $\langle W, R, V \rangle$ be a possible worlds model based on a transitive frame. Then

$$V(w) \models \Box\overline{A} \rightarrow \Box\overline{\Box\overline{A}}$$

for all $w \in W$ and $A \in \mathcal{L}_{\Box}$.

The first indication that possible worlds semantics for predicates yields something different is the realisation that a semantical version of Löb's theorem holds:

Proposition 3. For every possible world w in a possible worlds-model based on a transitive frame and for every sentence $A \in \mathcal{L}_{\Box}$, the following holds:

$$V(w) \models \Box\overline{\Box\overline{A} \rightarrow A} \rightarrow \Box\overline{A}.$$

The proof of this proposition proceeds like the proof of Löb's original theorem, using an instance of the diagonal lemma.

So far we have not even showed that there are possible worlds models in the sense described here. In operator modal logic, the existence of models is trivial, for one can turn the condition on V above into a recursive definition (recursion on the complexity of modal operator formulas). But our present predicate setting allows self-referential sentences, whereby valuation functions cannot be generally generated by a recursive definition. In fact, for many frames $\langle W, R \rangle$ there exists no model based on it. In general, when for a given frame there exists a model based on it, we say that the frame *admits* a model. It is then easy to generate examples of frames that do not admit a model.

Example 1. Consider the one-point reflexive frame, *i.e.*, the frame with $W = \{w\}$ and $R = \{\langle w, w \rangle\}$. This frame admits no model.

The proof of this is a semantical version of Montague's theorem:

Proof. Consider a diagonal sentence G such that $\mathsf{PA} \vdash G \leftrightarrow \neg\Box\overline{G}$. If $V(w) \models \neg G$, then by the diagonal property of G also $V(w) \models \Box\overline{G}$, whereby, since wRw, $V(w) \models G$. So $V(w) \models G$. Therefore, again by the diagonal property of G, there must be a world u such that $V(u) \models \neg G$. But since u must be identical with w, we have a contradiction.

A straightforward generalisation of the argument in this example shows that:

Proposition 4. No reflexive frame admits a model.

This suggests a question which has no direct counterpart in operator modal logic: *which frames admit models*?

Halbach, Leitgeb and Welch provide a partial solution to this question. First, they show that converse wellfounded frames admit exactly one valuation. This leaves open the question whether there exist converse ill-founded frames which admit valuations, and if so, how they look like. This question turns out to be nontrivial. The authors manage to give an informative answer, at least for the class of transitive frames.

In order to state their results, we first need some straightforward definitions related to the notion of wellfoundedness. Take any frame $\langle W, R \rangle$. We denote the *transitive closure* of R as R^*. And for every $w \in W$, we define $w\!\downarrow$ the set of all $v \in W$ such that wR^*v. Then we call a world $w \in W$ *converse wellfounded* if every subset of $w\!\downarrow$ has an R-maximal element. A set of worlds is wellfounded if all of its worlds are wellfounded. This then gives rise to the usual notion of converse wellfoundedness of frames that we have used without comment earlier: a frame $\langle W, R \rangle$ is converse wellfounded if and only if W is converse wellfounded. A *path* in a frame $\langle W, R \rangle$ below a world $w \in W$ is any maximal linearly ordered set of worlds in $w\!\downarrow$. We say that a world $w \in W$ has *height* α if α is the supremum of all order types of converse wellfounded paths below w. In this way, the notion of height of a world is defined even for converse illfounded worlds.

Two more definitions are needed. First, we define a second satisfaction relation $\models_{\Diamond}$:

Definition 1. $X \models_{\Diamond} A \equiv \langle \mathbb{N}, X^{\Diamond} \rangle \models A$,

i.e., $X \models_{\Diamond} A$ if and only if A is true if $\Diamond$ is interpreted by A. Second, we define the following operator $\Phi : X \mapsto \Phi(X)$ on $\wp(\mathbb{N})$:

Definition 2. $\overline{A} \in X \equiv (A \in X \text{ or } X \models_{\Diamond} A)$.

Φ is *not* monotone. Nevertheless it is not hard to see that it must have a least fixed point with an associated closure ordinal κ. If we generate a frame by applying Φ until we reach a least fixed point, then in the model associated with this frame the Reflexivity axiom will be true in the world associated with the closure ordinal κ.

In terms of this closure ordinal κ, a partial answer to the question which illfounded transitive frames admit models can be given:

Theorem 1. Assume $\langle W, R\rangle$ is transitive and every converse illfounded world in W has height at least κ. Then $\langle W, R\rangle$ admits a valuation.

The hard part is to determine is to determine the exact size of the closure ordinal κ. This ordinal turns out to be quite large, much larger than the least non-recursive ordinal, the Church-Kleene ordinal ω_1^{CK}. Halbach, Leitgeb and Welch show that κ is in fact the least level γ in the L-hierarchy of the constructible sets such that γ has a Σ_1-end extension. The proof of this uses deep facts from the theory of admissible sets.

The Halbach-Leitgeb-Welch approach admirably satisfies the Kripkean demand of providing “an area rich in formal structure and mathematical properties” (*cf.* [Kri$_1$75, p. 63]). Nevertheless, it also contains features which some will find unappealing. Many modal logicians will be disappointed that many of the systems of modal logic that they are familiar with (T, S4, S5) have no place in this set-up: the Reflexivity axiom can simply not be validated in all possible worlds. Also, some may find it undesirable that for converse wellfounded frames, the valuation function is forced upon us. But perhaps this must simply be seen as a consequence of the fact that the language for which the possible worlds semantics is given contains no nonmathematical vocabulary.

5 Axiomatic theories of intensional predicates

It is fair to say that with respect to axiomatic approaches to the logic of intensional predicates, the field is still wide open. One of the main problems is that, in contrast with the case of truth, we here have lacked semantic pictures guiding us.

We will discuss some attempts to develop axiomatic theories of knowability. In the context of the program of Epistemic Arithmetic, the notion of informal knowability has been investigated in an arithmetical context.[19] Knowability was there treated as a sentential operator; here we want to treat it as a predicate. However, some of the central philosophical heuristic ideas and research questions of the program of Epistemic Arithmetic prove to be fruitful also in the present predicate setting.

5.1 Predicate Epistemic Arithmetic

It appears natural to take the following crude and provisional (and, as we will see, ultimately unsatisfactory) diagnosis of the Paradox of the Knower as a starting point for developing an axiomatic theory of absolute knowability. Let us reconsider the inconsistent theory T_1. To which principle should we assign the blame for the contradiction? Reflexivity appears to be an immediate truth. It follows directly from an analysis of the concept of knowledge, knowledge being often thought to be *defined* as true justified belief (plus a Gettier condition, perhaps). The justification of the Provabilitation rule appears somehow less immediate. So we are inclined to assign the blame to Provabilitation. Suppose in addition that we want to adopt a principle of minimal mutilation. Then we will want to maintain as much of Provabilitation as we can have without bringing in the paradoxes. Closer analysis of the situation (for instance in the light of Thomason's argument) then shows that a crucial element in the derivation of the paradox is the Provabilitation of Reflexivity.[20] This then leads to the following heuristic principle:

Do Not Provabilitate Reflexivity

In other words, the hope is that by restricting Provabilitation so that Reflexivity is no longer in its range, the paradoxes can be avoided. This

[19] See the articles in [Sha$_1$85a]. Reinhardt has also contributed significantly to this research program. See, *e.g.*, [Rei$_0$86a].

[20] This is clearly brought out in [And83].

then leads to the formulation of a consistent system of Predicate Epistemic Arithmetic.

The basic system that is taken as a starting point is called PEA. This system is formulated in the language of Peano Arithmetic augmented with a new primitive predicate P. We call this language $\mathcal{L}_{\mathsf{PEA}}$. PEA consists of the closure under first-order logic of the axioms:

(1) PA in the extended language $\mathcal{L}_{\mathsf{PEA}}$, where occurrences of P are allowed in instances of the induction scheme;
(2) $\mathrm{P}\overline{A} \rightarrow A$ for all formulas $A \in \mathcal{L}_{\mathsf{PEA}}$;
(3) $\mathrm{Bew}_{\mathsf{BPEA}}(\overline{A}) \rightarrow \mathrm{P}\overline{A}$, for all formulas $A \in \mathcal{L}_{\mathsf{PEA}}$. Here $\mathrm{Bew}_{\mathsf{BPEA}}$ is the standard provability predicate for the **basis** BPEA of PEA,[21] which is the theory consisting of:

B1 PA in the extended language $\mathcal{L}_{\mathsf{PEA}}$, where occurrences of P are allowed in instances of the induction scheme;
B2 $\mathrm{P}\overline{A} \rightarrow (\mathrm{P}\overline{A \rightarrow B} \rightarrow \mathrm{P}\overline{B})$ for all formulas $A, B \in \mathcal{L}_{\mathsf{PEA}}$;
B3 $\mathrm{P}\overline{A} \rightarrow \mathrm{P}\overline{\mathrm{P}\overline{A}}$ for all formulas $A \in \mathcal{L}_{\mathsf{PEA}}$;

In this system, axiom 3 functions as a weakened version of the Provabilitation rule. It respects the injunction of the previous subsection not to provabilitate Reflexivity. Otherwise it is kept as strong as possible. B2 is (as before) called the *Distributivity axiom*. B3 is called the *4-axiom.*

PEA is the most straightforward implementation of the strategy outlined in the previous section. The first system that comes to mind when one wants to formalize the notion of informal provability is the modal system S4.[22] The Paradox of the Knower teaches us that if informal provability is treated as a *predicate* in the formalization, then the resulting system is inconsistent. If this inconsistent system is then slightly weakened by disallowing Reflexivity to be Provabilitated, what is obtained is precisely the system PEA. Slightly weaker variants of this system have been considered and proved to be consistent by several authors.[23]

[21] This way of formulating a formal theory of Predicate Epistemic Arithmetic is due to Friedman and Sheard in [FriShe87].

[22] History is our witness here. [Göd33] contains the first attempt to formulate an axiomatic theory of absolute provability (treated as an operator). He lists the S4 principles.

[23] See [$\mathrm{Ger}_1$70, p. 36-37]; [FriShe87, p. 7 (chart 1)]; [Nie91, p. 36]. PEA is considerably stronger than the system that was proposed in [Myh60, p. 469-470], which appears to be one of the earliest attempts to consistently formalise informal provability as a primitive predicate.

There are two ways in which one can contemplate strengthening PEA. Variants of PEA can be constructed by:

- adding a principle I to PEA, yielding a stronger system PEA + I;
- strengthening the basis of PEA, and modifying axiom 3. accordingly. For a given principle S, axiom 3. is replaced by

$$\mathrm{Bew}_{\mathsf{BPEA+S}}\left(\overline{A}\right) \rightarrow \mathrm{P}\overline{A}.$$

  The resulting system is then called PEA + S^i. The superscript i indicates that S is added to the 'inner logic' of the system. More on this below.

Of course hybrid strengthenings of PEA, of the form PEA + $\mathsf{S}^i + I$, can also be formed.

The following principles provide natural ways of strenthening PEA in one of the above two ways. First, there is the converse of the 4-principle:

$$\mathrm{P}\overline{\mathrm{P}\overline{A}} \rightarrow \mathrm{P}\overline{A},$$

which is called C4. Of course, C4 is *provable* in PEA. But we can contemplate whether it is consistent to include C4 even in the *basis* of PEA. For a second candidate additional axiom for PEA, consider the following principle of propositional (operator) modal logic, called *Fitch's Axiom*:

$$\Box\neg\Box A \rightarrow \Box\neg A.$$

It is well-known that in all extensions of the system T of propositional modal logic, this principle cannot be added without trivialising the modal operator.[24] But this argument depends crucially on an application of the Provabilitation rule to a sentence obtained by an application of the Reflexivity axiom. The corresponding argument in PEA would therefore break down: we would not be allowed to provabilitate. Therefore the possibility arises that the 'predicate counterpart' F:

$$\mathrm{P}\overline{\neg\mathrm{P}\overline{A}} \rightarrow \mathrm{P}\overline{\neg A}$$

[24] This observation is due to Fitch, *cf.* [Fit$_0$63]. In the philosophical community, Fitch's argument has generated an extensive discussion. For an overview, see, *e.g.*, [Wil$_1$00, Chapter 12].

of Fitch's Axiom can be added to PEA without trivialising the informal provability predicate. Thirdly, the converse of F (CF) can be considered:

$$\mathrm{P}\overline{\neg A} \rightarrow \mathrm{P}\overline{\neg \mathrm{P}\overline{A}}.$$

In all extensions of the operator modal logic S4, (the operator part of) CF is derivable. But again this derivation makes use of the ability to provabilitate sentences obtained by Reflexivity, the counterpart of which we do not have in PEA.

Note that the soundness of the Barcan Formula (BF)

$$\mathrm{P}\overline{\exists x A} \rightarrow \exists y \mathrm{P}\overline{A\left[\dot{y}\,/x\right]}$$

appears highly questionable for the intended interpretation of P. Therefore BF will not be considered as a putative extra axiom. A consequence of this is that it is not to be expected that the extensions of PEA will be arithmetically significantly stronger than the arithmetical systems on which they are based. For (as Sheard observed), BF generally appears to play a crucial role in the added arithmetical strength of axiomatic systems of truth.

5.2 Consistency questions and results

It seems that not all the consistency questions that these proposals raise are equally easy to settle. There are the following results:[25]

Theorem 2. PEA + CF^{i} has an ω-model.

This theorem entails of course that:

Corollary 1. PEA has an ω-model.

We recall the definition of the notion of *uniform reflection* for PA:

Definition 3. URfn(PA) is the scheme:

$$\mathrm{Bew}_{\mathsf{PA}}\left(\overline{A}\right) \rightarrow A \text{ for all formulas } A \in \mathcal{L}_{\mathsf{PA}}.$$

Then we have:

[25] For proofs and a more detailed discussion of these results, see [Hor$_1$02].

Theorem 3. For all formulas $A \in \mathcal{L}_{PEA}$:

$$\mathsf{PEA} + F^i \vdash A \Leftrightarrow \mathsf{PA} + \mathsf{URfn}(\mathsf{PA}) \vdash A.$$

Corollary 2. $\mathsf{PEA} + F^i$ is consistent.

In view of these results, one might suspect that as long as we do not provabilitate reflexivity, we can have it all, *i.e.*, we can consistently add F, CF, 4, C4 simultaneously to the basis of PEA. But this is not true. One can show, as a consequence of a theorem of Friedman and Sheard, that already:

Proposition 5. (Leitgeb) $\mathsf{PEA} + \mathsf{C4}^i$ is inconsistent.

The following problem is, as far as I am aware, still unresolved:

QUESTION: Is $\mathsf{PEA} + \mathsf{F}^i + \mathsf{CF}^i$ consistent?

5.3 Inner versus outer logic

There exists an analogy between certain axiomatic theories of truth and the systems of Predicate Epistemic Aritmetic that we are considering. Reinhardt has noted that the system KF is only *partially* sound for the notion of truth.[26] Let L be the liar sentence. Then KF proves *both* L and $\neg \mathrm{T}\overline{L}$. In other words, some sentences that are proved by KF are *denied truth* by KF itself. So we cannot believe everything that KF proves.[27]

A similar phenomenon occurs for PEA and its relatives. Consider again the absolute Gödel sentence G. In the 'weak' system PEA, G is easily seen to be provable. But at the same time, PEA denies G to be informally provable. In other words, PEA explicitly denies that all its derivations are 'honest to God' proofs.

Reinhardt's proposed solution to this problem was to restrict the attention to the *inner logic* of the respective systems. For instance, for KF, we should restrict our attention to the (in principle axiomatizable) subsystem consisting of those sentences A such that $\mathsf{KF} \vdash T\overline{A}$. In a similar vein, presumably, in the case of PEA the proposal would be to restrict our attention to the sentences A such that $\mathsf{PEA} \vdash \mathrm{P}\overline{A}$.

Let IPEA designate the inner logic for PEA, etc. Here are the formalizations of the inner logics of some of the variants of PEA that we have been considering:

---

[26] *Cf.* [Rei$_0$86a].

[27] Something similar holds for Cantini's system VF.

– $\mathsf{IPEA} \equiv \mathsf{PA} + A \Rightarrow \mathrm{P}\overline{A} + \mathrm{P}\overline{A} \rightarrow (\mathrm{P}\overline{A \rightarrow B} \rightarrow \mathrm{P}\overline{B}) + \mathrm{P}\overline{A} \rightarrow \mathrm{P}\overline{\mathrm{P}\overline{A}}$
– $I\,(\mathsf{PEA} + \mathsf{F}^{i}) \equiv \mathsf{IPEA} + \mathrm{P}\overline{\neg \mathrm{P}\overline{A}} \rightarrow \mathrm{P}\overline{\neg A}$[28]

Whereas PEA and its extension cannot be regarded as *sound* systems for intuitive provability, their inner logics IPEA,... *are* sound formalizations of the intuitive notion of provability. It is hardly necessary to mention that these inner logics are *significantly weaker* than the 'outer logics' from which they are derived.

But this means that the situation is quite different from the way it was made to appear in Section 5.1. There it was suggested that essentially the only price that needs to be paid for avoiding the Paradox of the Knower is to refrain from provabilitating Reflexivity. Now we see that even if Reflexivity is only assumed *as a hypothesis*, so to speak, pathological results follow. When we focus on the inner logic of PEA, we see that Reflexivity is *not* contained in it —but it *does* contain full Provabilitation![29] So the blame for the Paradox of the Knower is now shifted from Provabilitation to Reflexivity. We have given up our faith in full Reflexivity altogether.

One may around this point start to wonder why we even bother to investigate these outer logics.[30] There is at least one good reason. *If* we find an ω-model $\mathfrak{I}$ for an outer logic containing Reflexivity, then we immediately have a *sound* model for the associated inner logic, in the precise sense that for all sentences A, if $\overline{A}$ belongs to the extension of P according to $\mathfrak{I}$, then $\mathfrak{I} \models A$. There may be additional reasons. Perhaps an argument can somehow be given that we are entitled to a *larger* part of the outer logic than what is contained in the corresponding inner logic? But this is at this point just idle speculation.

5.4 Church's Thesis and relations with intuitionistic arithmetic

One of the attractive features of operator Epistemic Arithmetic is that in these systems epistemic analogues of the *Church-Turing Thesis* (CT) can

[28] This raises the question whether the system

$$\mathsf{IPEA} + \mathrm{P}\overline{\neg \mathrm{P}\overline{A}} \rightarrow \mathrm{P}\overline{\neg A} + \mathrm{P}\overline{\neg A} \rightarrow \mathrm{P}\overline{\neg \mathrm{P}\overline{A}}$$

is consistent. This question is related to the open problem presented at the end of Section 5.2.

[29] In this respect, we have now argued ourselves into a position that is in line with the Halbach-Leitgeb-Welch approach.

[30] Note that the same question can be asked about many popular axiomatic theories of truth. In my opinion, logical and philosophical questions regarding the relation between inner logic and outer logic have until now not received the attention they deserve.

be expressed and that under Gödel's translation intuitionistic arguments can be formalized in a classical (but epistemic) context. It is natural to ask to what extent this can also be done in Predicate Epistemic Arithmetic. The results that are known in this context are the following.[31]

First, consider the following predicate epistemic analogue (ECT) of CT:

$$\overline{\mathrm{P}\forall x\exists y\mathrm{P}\overline{A\left(\dot{x},\dot{y}\right)}} \rightarrow \exists e\forall x\exists y\left(\mathrm{T}\left(e,x,y\right)\wedge A\left(x,U\left(y\right)\right)\right),$$

where T is Kleene's T-predicate and U is the U-function symbol. This principle is interesting, for it is the sole principle proposed to date which on the one hand seems plausible when P is interpreted as knowability, but on the other hand is definitely implausible when P is interpreted as truth. It can be shown that PEA + ECT is consistent. But unfortunately ECT cannot be consistently added to the inner logic of PEA.[32]

QUESTION: Can the restriction of the scheme ECT to *nonepistemic* predicates $A(x, y)$ be consistently added to the inner logic of PEA?

Second, for PEA (and several of its consistent natural extensions) the following epistemic analogues of the disjunction property and the numerical existence property for intuitionistic arithmetical systems can be shown to hold:

Theorem 4. :

- (ENEP) PEA $\vdash \mathrm{P}\overline{A} \vee \mathrm{P}\overline{B} \Rightarrow$ PEA $\vdash \mathrm{P}\overline{A}$ or PEA $\vdash \mathrm{P}\overline{B}$ for all sentences $A, B \in \mathcal{L}_{\mathsf{PEA}}$.
- (EDP) PEA $\vdash \exists x\mathrm{P}\overline{A} \Rightarrow$ PEA $\vdash \mathrm{P}\overline{A\left[\underline{n}/x\right]}$ for some $n \in \mathbb{N}$ and for all formulas $A \in \mathcal{L}_{\mathsf{PEA}}$.

Thirdly, in operator Epistemic Arithmetic a natural extension g of Gödel's translation from intuitionistic logic to S4 modal logic[33] has been considered. It was proved that g is a *faithful* translation from Heyting Arithmetic (HA) to Shapiro's system EA of Epistemic Arithmetic:[34]

---

[31] For proofs of these results, the reader is again referred to [Hor$_1$02].
[32] This observation is due to Halbach.
[33] *Cf.* [Göd33].
[34] An elegant proof of this theorem is given in [FriFla86].

Theorem 5. For all sentences A of the language of intuitionistic arithmetic:

$$\mathsf{HA} \vdash A \Leftrightarrow \mathsf{EA} \vdash g(A).$$

For the natural counterpart $g' : \mathcal{L}_{HA} \mapsto \mathcal{L}_{PEA}$ in our predicate setting of this translation $g : \mathcal{L}_{HA} \mapsto \mathcal{L}_{EA}$, the corresponding faithfulness theorem to all appearences does not hold. Already for the soundness direction of this theorem, it would appear that we need to have C4 to the inner logic of PEA, which we know to be impossible.

6 Truth and Intensionality

The paradoxes about knowledge, necessity, truth and rational belief are produced by very similar diagonal arguments. In this sense, at least, these paradoxes are closely related. Let us now turn to the philosophical question that was left hitherto unaddressed: *why* is it that the notions of necessity, informal provability, and perhaps rational belief are subject to liar-like paradoxes?

There is a view which holds that the relation between these paradoxes goes deeper. This view appears to have numerous adherents. But it is rarely made explicit or defended in the literature. The idea is that there is an underlying *conceptual connection* between the notions involved which explains *why* they are all paradoxical. This view is usually combined with the belief that at bottom there is only *one* paradoxical concept: truth—or even a bit more fundamentally, satisfaction. All the liar-like paradoxes are just manifestations of the paradoxicality of the concept of truth. But one wonders exactly in which way this is so. For in the derivation of the liar-like paradoxes, the concept of truth does not occur.

Perhaps a clue to the solution of this problem can be taken from possible worlds semantics. Instead of formalizing truth as a one-place predicate $T(x)$, perhaps we ought really to formalize it as a *many-place* predicate $T(x, w, t, i)$, which should be read as "x is true in possible world w at time t and place i".

If this is on the right track, then the necessity predicate might be parsed as:[35]

$$\forall w \forall t \forall i T(x, w, t, i).$$

This would mean that Montague's Paradox involves the truth predicate after all. And since the notion of intuitive prova*bility* or knowa*bility* implicitly contains a modal component, it too would indirectly contain the concept of truth. Therefore the Paradox of the Knower would also involve the truth predicate.[36] Moreover, one would be lead to suspect that temporal and spatial notions might also be subject to liar-like paradoxes, which we have indeed seen to be the case. Thomason's paradox about rational belief remains. It is indeed not easy to see in what sense the notion of rational belief contains a modal, temporal, or spatial indexical component. But if what I have said in Section 2 about this argument is correct, then it is not clear that we *have* a genuine paradox on our hands here.

In sum, the suggestion that is tentatively put forward here is the following. In exclusively looking at the syntactical and semantical techniques and results concerning truth and the semantic paradoxes for ideas for analysing necessity, knowability and temporal notions as predicates, we may to some extent have approached the matter from the wrong end. Perhaps it is time to look at the 'possible worlds'-tradition in modal logic for guiding us in our logical analysis of the notion of truth.

[35] If one takes a Tarskian point of view, then one will insist on one more parameter: one will insist that truth is also always relative to a *language*.

[36] Note, in this context, that the notion of *actually having* an informal proof is not, as far as we know, subject to liar-like paradoxes.