POSSIBLE-WORLDS SEMANTICS FOR MODAL NOTIONS CONCEIVED AS PREDICATES

VOLKER HALBACH, HANNES LEITGEB, AND PHILIP WELCH

Abstract. If $\Box$ is conceived as an operator, i.e., an expression that gives applied to a formula another formula, the expressive power of the language is severely restricted when compared to a language where $\Box$ is conceived as a predicate, i.e., an expression that yields a formula if it is applied to a term. This consideration favours the predicate approach. The predicate view, however, is threatened mainly by two problems: Some obvious predicate systems are inconsistent, and possible-worlds semantics for predicates of sentences has not been developed very far. By introducing possible-worlds semantics for the language of arithmetic plus the unary predicate $\Box$, we tackle both problems. Given a frame $\langle W, R\rangle$ consisting of a set W of worlds and a binary relation R on W, we investigate whether we can interpret $\Box$ at every world in such a way that $\Box\ulcorner A\urcorner$ holds at a world $w \in W$ if and only if A holds at every world $v \in W$ such that wRv. The arithmetical vocabulary is interpreted by the standard model at every world. Several 'paradoxes' (like Montague's Theorem, Gödel's Second Incompleteness Theorem, McGee's Theorem on the ω-inconsistency of certain truth theories etc.) show that many frames, e.g., reflexive frames, do not allow for such an interpretation. We present sufficient and necessary conditions for the existence of a suitable interpretation of $\Box$ at any world. Sound and complete semi-formal systems, corresponding to the modal systems K and $\mathsf{K4}$, for the class of all possible-worlds models for predicates and all transitive possible-worlds models are presented. We apply our account also to nonstandard models of arithmetic and other languages than the language of arithmetic.

Received by the editors October 9, 2002.

Key words and phrases. Possible worlds, modal logic, necessity, paradox.

We have presented parts of the paper in several talks; some of the results are published in [19] without proofs. We thank the audiences and the referee of the latter paper for numerous suggestions and hints. We are also indebted to Christopher von Bülow for his suggestions and for correcting many typographical errors. The work of Volker Halbach was sponsored by the Department of Logic and Philosophy at the University of California at Irvine and the research group *Logic in Philosophy* at the University of Constance. Philip Welch would like to acknowledge the support of the Institut für Formale Logik, Vienna, where he held a Guest Professorship during the preparation of part of the paper.

> [...] semper denotatur quod talis modus verificetur de propositione illius dicti, sicut per istam 'omnem hominem esse animal est necessarium' denotatur quod iste modus 'necessarium' verificetur de ista propositione 'omnis homo est animal' cuius dictum est hoc quod dicitur 'omnem hominem esse animal'[.]
>
> William of Ockham, Summa Logicae II.9

§1. Predicates and Operators. Predicates applied to singular terms yield formulae, while operators need to be combined with formulae to give new formulae. Roughly speaking, in natural language and in the case of necessity "necessarily" and "it is necessary that" are operators, whereas "is necessary" is a predicate. Whether necessity, knowledge, belief, future and past truth, obligation and other modalitities should be formalised by operators or by predicates was a matter of dispute up to the early sixties between two almost equally strong parties. Then two technical achievements helped the operator approach to an almost complete triumph over the predicate approach that had been advocated by illustrious philosophers like Quine.

Montague [33] provided the first result by proving that the predicate version of the modal system T is inconsistent if it is combined with weak systems of arithmetic. From his result he concluded that "virtually all of modal logic ... must be sacrificed", if necessity is conceived as a predicate of sentences. Of course, Montague's verdict does not imply that necessity cannot be treated as a predicate of objects different from sentences, e.g., propositions conceived as language-independent entities, but the result clearly restricted the attractiveness of the predicate approach.

The majority of philosophers and logicians have opted to spare modal logic and to reject the predicate view altogether. The prevailing general attitude is expressed by the following quote from Slater [43]:

> Since Montague, we surely now know that syntactic treatments of modality must be replaced by operator formulations.

The other technical achievement that brought about the triumph of the operator view was the emergence of possible-worlds semantics. Hintikka, Kanger and Kripke provided semantics for modal operator logics, while nothing similar seemed available for the predicate approach.

Ever since possible-worlds semantics has reigned as the main tool in the analysis of necessity and the notions mentioned above. Today sacrificing possible-worlds semantics for the predicate conception of necessity etc. would mean sacrificing a huge body of philosophical logic and of analytic metaphysics, epistemology, ethics and computer science.

Nevertheless the operator approach suffers from a severe drawback: it restricts the expressive power of the language in a dramatic way because it rules out quantification in the following sense: There is no direct formalisation of a sentence like "All tautologies of propositional logic are necessary".

Proponents of the operator approach have proposed several strategies to overcome this problem. For instance, one might formalise the above sentence by a scheme and transfer quantification into the metalanguage. However, this strategy hardly will be satisfying if it is applied to more complex sentences where the quantifiers are deeply embedded in the sentence.

Alternatively, one might employ a special kind of quantification where the above sentence can be formalised as $\forall A\,(P(A) \to \Box A)$, where $P(x)$ is a predicate while $\Box$ still is an operator. Then even $\forall A\,(P(A) \to A)$ would be well formed because the variable A can stand in place of a formula. This kind of quantification often is called *substitutional* quantification, sometimes even if it does not come with any semantics but only with axioms and rules. This terminology is surely confusing, but it has become almost standard terminology after Kripke [27] had shown how to apply a substitutional interpretation of quantification with the mentioned properties. Substitutional quantification in this sense is not simple propositional quantification because the new variables may appear in the place of normal first-order variables as well.

It seems arbitrary that some notions are formalised as predicates, while others are conceived as operators only. It also forces one to switch between usual first-order quantification and substitutional quantification without any real need. In natural language there are not these two different kinds of quantification. We shall give an example. "All true Σ-sentences are provable in PA" requires only a first-order quantifier because provability and truth are predicates. "All true Σ-sentences are provable in PA and necessary", in contrast, is formalised with another kind of quantification. But it is completely unclear why we should assume that in the latter sentence we deal with a completely different kind of quantification than in the first sentence.

In most cases the above kind of substitutional quantification will be equivalent to a truth predicate. If a truth predicate is employed, then one could stick to an operator for necessity in many cases. Instead of saying "x is necessary", one could say "x is necessarily true". Thus $\Box x$ where $\Box$ is a predicate might be replaced by $\Box \mathrm{T}x$ where $\Box$ is an operator and T is a truth predicate. One could even reverse the order of the truth predicate and the necessity operator (see Kripke [26]). Thus only a truth predicate (or substitutional quantification) would be needed, but otherwise operators would suffice.

We have several qualms about this approach. As above, we do not see any good reason to treat truth and necessity (and the other predicates) differently on the syntactical level. Furthermore the theory of necessity and other notions would rest on truth-theoretic foundations which are threatened by the semantical paradoxes. In general, the theory of truth is far from being settled and the theory of necessity would inherit all semantical paradoxes.

We think that the operator approach might have some merits for instance in linguistics and computer science, but it fails at its main application in philosophical logic: it does not provide an illuminating analysis for necessity, knowledge, obligation and so on. For it does not allow for the formalisation of the most common philosophical claims such as "All laws of physics are necessary" or "There are true but unknowable sentences".

Of course, the rejection of substitutional quantification and the other strategies mentioned requires a more detailed discussion. However, we do not present it here but proceed to our own constructive proposals.

Even for those who adhere to the operator view our results might be interesting because the difficulties we encounter will show up in a different form if a proper account of substitutional quantification or a theory of truth is used in order to express quantification.

Our general reservations against the operator approach are shared by some philosophers (see also Bealer [4] for an overview). Thus along the main stream of the operator conception with its possible-worlds semantics there always has been a tiny but steady rill of work on the predicate view by authors like Germano [15], Skyrms [42], Asher and Kamp [2], McGee [32], Schweizer [41], Gupta and Belnap [5, chapter 6].

Some of these papers have unmasked several of the alleged arguments against the predicate view as untenable prejudices. In the first place, everything that can consistently be said in an operator language can be consistently said in the language with a predicate instead. More precisely, sentences of modal operator systems can be translated into sentences of a rich predicate system in a natural way such that provability is preserved. The resulting system with $\Box$ as a predicate is consistent with systems that allow for a proof of the Diagonalisation Theorem. Under this translation, operator sentences become sentences where the modal operator is replaced by a corresponding predicate. This argument can be taken as a proof that one does not have to sacrifice anything for the predicate view (see des Rivières and Levesque [38] and Gupta and Belnap [5, p. 240 ff]).

Only if the provability of axiom *schemata* is to be preserved, troubles are emerging: we can have $\Box\ulcorner A\urcorner \to A$ for any translation of a modal operator formula into the language with $\Box$ as a predicate, but we cannot have it for all sentences of the predicate language. The translations of operator sentences are always 'grounded' sentences, that is, roughly speaking, sentences whose semantical evaluation hinges on nonsemantical facts; more exactly, they are sentences that arise from sentences without $\Box$ by finite applications of $\Box$. Sentences like the liar sentence that yield the inconsistencies are typical examples of ungrounded sentences; they do not have counterparts in the operator language.

In this paper we investigate the second main argument against the predicate approach which says that possible-worlds semantics is not feasible for predicates. Asher and Kamp [2] and Gupta and Belnap [5, chapter 6] have already refuted this view by providing possible-worlds semantics for modal predicate languages in certain special cases. But although in some cases possible-worlds semantics can be provided, there are also well-known severe restrictions. We want to explore in which cases problems arise.

We shall stick as closely as possible to the common possible-worlds semantics for operators, and we shall not apply tricks to avoid paradoxes. That is, we shall use classical logic throughout, in contrast to other authors who have applied the techniques known from the theory of truth, e.g., partial or many-valued logic, to the theory of necessity and similar concepts.

However, there is one fundamental decision we have to make which is avoided on the operator approach. If necessity, truth, belief and the like are predicates, then they must apply to objects of a certain kind. Propositions, tokens or types of sentences, utterances, mental objects or the like have been used for this purpose. The different choices have been discussed for a long time. Here we are not able to go into this discussion and mention just one aspect.

The problem of the choice between the different categories of objects is connected to a family of arguments that have been employed against necessity, truth and belief as predicates of sentences (e.g., the so-called Church–Langford argument; see [4]). According to these arguments, the meaning of a linguistic expression is always contingent. Thus, for instance, it is never necessary that "All men are men" is necessary, because the expression "All men are men" could have meant something else; it could have meant "It is raining". Therefore the expression "All men are men" is not necessarily necessary; it is only contingently necessary, namely contingently on the assumption that the string of symbols has its actual meaning. Only the *proposition* expressed by "All men are men" is necessarily necessary.

Arguments of this kind have lead philosophers to the claim that truth, belief etc. are predicates of propositions (e.g., in the case of Bealer and Pap [36]), while others have claimed that truth is not a predicate at all (e.g., Strawson [44] and Grover, Camp & Belnap [16]). Still others (like Field [13]) maintain that there are 'deflationist' uses of the truth predicate to which this kind of reasoning simply does not apply. We do not discuss these arguments here.

It is compatible with our account to conceive $\Box$ either as a predicate of sentences or as a predicate of propositions—as long as the latter share the structure of sentences. If we opt for sentences as objects to which necessity, truth etc. is ascribed, then we have to say that we deal with 'deflationist' uses of these predicates. At any rate we keep the interpretation of the vocabulary fixed that talks about the objects to which truth, necessity and so on is ascribed. So the argument involving the contingency of meanings is avoided.

In our technical treatment we shall apply the predicate $\Box$ to numerical codes of sentences (and we shall even identify expressions with their codes), although we do not think that this is actually the most satisfying approach. It allows us, however, to proceed to our main topic, possible-worlds semantics for predicates, without before developing a theory of propositions or other objects.

Our approach leaves room for several interpretations: the codes could be interpreted as types of sentences or as (language-independent) propositions. We shall discuss these issues in a separate paper and return to our main topic.

The core of possible-worlds semantics is the analysis of necessity (or the notions mentioned above) as truth in all accessible worlds (or situations or the like). That is, sentences $\Box A$ (on the operator account) or $\Box\ulcorner A\urcorner$ (on the predicate account) are true if and only if A is true in all accessible worlds. Here a fundamental difference shows up between possible-worlds semantics for modal operator and modal predicate languages. In the case of an operator conception, this analysis can be turned into a clause in the definition of validity of modal sentences. If a frame, i.e., a set of worlds with an accessibility relation, is given and a function that assigns a model for the $\Box$-free language to any world, then it is easy to define truth at a world for all sentences with $\Box$ as an operator. The definition proceeds by recursion on the complexity of the sentence. Thus one can build a model on every given frame.

If $\Box$ is treated as a predicate, the recursive definition of truth at a world cannot be carried out anymore, because a sentence $\Box\ulcorner A\urcorner$ is atomic, while on the operator account $\Box A$ has the complexity of A plus one. Also there is no way to provide an amended definition of the complexity of a formula that allows for a recursive definition of truth at a world. Problems are posed, in particular, by self-referential sentences like the liar sentence.

For some frames a recursive definition of truth at a world is not only impossible, but there is not even a model based on the frame at all. Montague's [33] Theorem, for instance, shows that one cannot build possible-worlds models on reflexive frames (see Example 5 below) if necessity is viewed as a predicate. This means that there is no way to assign truth and falsity to all sentences in such a way that $\Box\ulcorner A\urcorner$ comes out as true if and only if A holds at all accessible worlds (and if some conditions are met which will be explained below). Another example is provided by McGee's [31] Theorem on ω-inconsistent systems. It is a direct consequence of this theorem that a frame consisting in the natural numbers ordered by the successor relation (so that every world sees exactly one other world) does not support a possible-worlds model.

Although several further such restrictions on frames are known, the general problem has not been tackled so far: which frames support a possible-worlds model, if $\Box$ is a predicate?

A satisfying answer to this characterisation problem would exhibit a graph-theoretical description of the frames supporting possible-worlds models. The restriction imposed by theorems like those due to Montague and McGee should drop out as consequences of such a characterisation result.

Our development of possible-worlds semantics for predicates will also yield some further insights. It allows to model certain paradoxes and inconsistency results in a framework where are modalities are treated uniformly. Here we explicitly reckon—like Ockham did—-also truth, being known and other notions among the modalities.

Moreover our approach sheds a new light on hierarchies of metalanguages, and some results of provability logic and modal logic are generalised. Our treatment of transitive converse wellfounded frames can be seen as a variant or generalisation of the revision theory of truth.

Overview of the remaining sections. In section 2 we define the basic notions of possible-worlds semantics such as frames etc. and apply this account to predicates rather than operators of sentences. Some examples of frames that do not support a model are provided in section 3. We tackle the Characterisation Problem for transitive frames in sections 4 and 5, and provide some results on other frames in section 6. In section 7 we show that some frames support more than one model even in the absence of "contingent" vocabulary. We return to a classical topic of operator modal logic in section 8, namely completeness; we shall provide a completeness result by the method of canonical models. Sections 9 and 10 deal with the properties of a fixed point of an operator that arises in the construction of models. All our constructions so far are based on the standard model of arithmetic; we sketch some applications to nonstandard models and other models and languages in sections 11 and 12.

§2. Possible-Worlds Semantics. $\mathcal{L}_{\mathrm{PA}}$ is the language of arithmetic; $\mathcal{L}_{\Box}$ is $\mathcal{L}_{\mathrm{PA}}$ augmented by the one-place predicate symbol $\Box$. We shall identify $\mathcal{L}_{\mathrm{PA}}$ with the set of all arithmetical formulae and $\mathcal{L}_{\Box}$ with the set of all formulae that may contain $\Box$. We do not distinguish between expressions (like formulae and terms) and their arithmetical codes. We use uppercase roman letters A, B, ... for sentences of $\mathcal{L}_\Box$. If a formula may have a free variable we always indicate this by writing $A(v)$ etc.

We shall assume that the language $\mathcal{L}_{\mathrm{PA}}$ features certain function symbols. This will render the notation somewhat more perspicuous. Of course, these function symbols can be eliminated in the usual way and thus are not really required.

On our account all arithmetical truths come out necessary. Thus one might wish to include further vocabulary that allows for more contingent sentences. In fact, our account can and should be applied to expansions of $\mathcal{L}_{\mathrm{PA}}$. In the main parts of this paper, however, we shall concentrate on the language $\mathcal{L}_{\mathrm{PA}}$ because additional vocabulary would be idling and obstruct the view for the essentials. We shall discuss expansions of $\mathcal{L}_{\mathrm{PA}}$ in section 12.

Except for section 11 we shall consider *standard* models only, i.e., models with the set of natural numbers as the domain and the standard interpretation of the arithmetical vocabulary. Hence models have the form $(\mathbb{N}, X)$ where $\mathbb{N}$ is the standard model of arithmetic and $X \subseteq \omega$ the extension for $\Box$. Since models differ only in the extension of $\Box$, we may identify a model with an extension for $\Box$. Thus we write $X \models A$ for $(\mathbb{N}, X) \models A$.

$\Diamond x$ is defined as $\neg\Box\dot{\neg}x$ where $\dot{\neg}$ is a function symbol for the function sending the code of a sentence of $\mathcal{L}_\Box$ to the code of its negation.

The definition of frames and possible-worlds models (PW-models, for short) parallels the usual definitions for the operator approach.

Definition 1. A *frame* is an ordered pair $\langle W, R\rangle$ where W is nonempty and R is a binary relation on W.

The elements of W are called *worlds*, R is the *accessibility* relation. World w *sees* v if and only if wRv.

Next we define PW-models. They must not be confused with the models of the form $(\mathbb{N}, X)$. The latter are models in the usual sense and act, loosely speaking, as worlds in PW-models.

Definition 2. A *PW-model* is a triple $\langle W, R, V\rangle$ such that $\langle W, R\rangle$ is a frame and V assigns to every $w \in W$ a subset of $\mathcal{L}_\Box$ such that the following condition holds:

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

If $\langle W, R, V\rangle$ is a model, we say that the frame $\langle W, R\rangle$ *supports* the PW-model $\langle W, R, V\rangle$ or that $\langle W, R, V\rangle$ *is based on* $\langle W, R\rangle$. A frame $\langle W, R\rangle$ *admits a valuation* if there is a valuation V such that $\langle W, R, V\rangle$ is a PW-model. Thus a frame supports a model if and only if it admits a valuation.

V assigns to every world a set of sentences, the extension of $\Box$ at that world. The condition on the function V in the definition says that a sentence is in the extension of $\Box$ at a world w if and only if it is true in all worlds seen by w. Thus if $\langle W, R, V\rangle$ is a PW-model, the following holds:

$$V(w) \models \Box\ulcorner A\urcorner \quad \text{iff} \quad \forall v \in W(wRv \;\Rightarrow\; V(v) \models A).$$

If we were to include further 'contingent' vocabulary, the valuation V not only would have to interpret $\Box$ but also this additional vocabulary. We shall consider this situation in section 12.

As pointed out above, the definition of truth at a world cannot be carried out recursively as in the operator case. On the operator account one can build models on every frame. Thus the question whether a frame supports a model does not arise for $\Box$ as an operator. However, if $\Box$ is a predicate, there are restrictions and the following question is sensible:

CHARACTERISATION PROBLEM. *Which frames support PW-models?*

An answer to this problem exhibits the restrictions imposed by the 'paradoxes' (like Montague's) on possible-worlds semantics for predicates. We are only able to provide partial solutions to the Characterisation Problem in this paper.

We shall now show that possible-worlds semantics for predicates is in many respect similar to the usual possible-worlds semantics for operators.

Analogous definitions of frames and models for operators lead to the so-called *normal* systems of modal logic with the minimal system K (see, e.g., Boolos [6]). These systems are closed under necessitation, and the necessity operator $\Box$ distributes over material implication. For the predicate account something similar can be shown.

LEMMA 3 (Normality). *Suppose $\langle W, R, V\rangle$ is a PW-model, $w \in W$ and $A, B \in \mathcal{L}_\Box$. Then the following holds:*

(i) If $V(u) \models A$ for all $u \in W$, then $V(w) \models \Box\ulcorner A\urcorner$;
(ii) $V(w) \models \Box\ulcorner A \to B\urcorner \to (\Box\ulcorner A\urcorner \to \Box\ulcorner B\urcorner)$.

Since we are dealing with standard models only, we do not only obtain schemata like (ii) but also their universal closure, that is

$$\forall x\, \forall y\, \Big(\mathrm{Sent}_{\mathcal{L}_\Box}(x) \wedge \mathrm{Sent}_{\mathcal{L}_\Box}(y) \to \big(\Box(x \dot{\to} y) \to (\Box x \to \Box y)\big)\Big).$$

Here $\dot{\to}$ represents the function that yields, when applied to two codes of sentences, the code of their material implication (if no such function symbol is available it is to be expressed by a suitable predicate expression). $\mathrm{Sent}_{\mathcal{L}_\Box}(x)$ represents the set of sentences of $\mathcal{L}_\Box$. In order to avoid confusing notation, we shall state schemata like (ii) instead of the universally quantified sentences. This does not make a difference because of the use of standard models.

§3. Some Limitative Results. As has been noted in the introduction, several restrictions on the class of frames supporting a PW-model are well known. These restrictions correspond to several inconsistency results. We begin with the most trivial example, namely the liar paradox or Tarski's Theorem which shows that object- and metalanguage cannot coincide.

EXAMPLE 4 (Tarski). The above frame with one world that sees itself does not admit a valuation.

Montague's Theorem is a generalisation thereof.

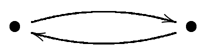
EXAMPLE 5 (Montague). If $\langle W, R\rangle$ admits a valuation, then $\langle W, R\rangle$ is not reflexive.

The trivial proof resembles Montague's proof showing that the predicate version of the modal system T is inconsistent. We present it in detail as an example.

PROOF. Assume $\langle W, R, V\rangle$ is a PW-model based on $\langle W, R\rangle$ which is reflexive. The liar sentence is a sentence such that $\mathrm{PA} \vdash A \leftrightarrow \neg\Box\ulcorner A\urcorner$; it holds at any world. Pick an arbitrary world $w \in W$.

If $V(w) \models \neg A$ then also $V(w) \models \Box\ulcorner A\urcorner$ and thus by reflexivity $V(w) \models A$. Therefore $V(w) \models A$ must hold and consequently there is a world u such that wRu and $V(u) \models \neg A$ and again $V(u) \models A$. $\dashv$

The arguments showing that a frame does not admit a valuation proceed by diagonal arguments. In the above examples we employed the liar sentence. The next two examples require slightly more complex diagonal sentences.

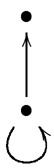


EXAMPLE 6. The frame 'two worlds see each other' displayed above does not admit a valuation.

The existence of a suitable valuation is refuted by the fixed point $A \leftrightarrow \neg\Box\ulcorner\Box\ulcorner A\urcorner\urcorner$.

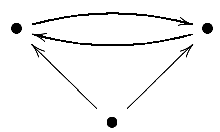
This example can be generalised in the obvious way in order to refute the existence of valuations for loops of arbitrary finite length.

Now we turn to a frame with a dead end, i.e., a world which does not see any other world.



Example 7. The frame 'one world sees itself and one other world' does not admit a valuation.

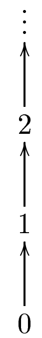
For the proof the fixed point $A \leftrightarrow (\Box\ulcorner A\urcorner \rightarrow \Box\ulcorner\neg A\urcorner)$ can be employed.



Example 8. The above frame 'one world sees two worlds that see each other' does not admit a valuation.

One can show this by using the fixed point $A \leftrightarrow \neg\Box\ulcorner\Box\ulcorner A\urcorner\urcorner \wedge \neg\Box\ulcorner A\urcorner$.

So far we have been dealing with transitive frames only. McGee's main theorem in [31] on the ω-inconsistency of a certain theory of truth imposes a restriction on the existence of PW-models based on a nontransitive frame. Pre is the predecessor relation: $k \,\mathsf{Pre}\, n$ if and only if $n = k + 1$.



Example 9 (McGee, Visser). $\langle\omega, \mathsf{Pre}\rangle$ does not support a PW-model.

The proof relies on the following diagonal sentence involving a quantifier (see McGee [31]):

$$A \leftrightarrow \neg\forall x\, \Box \underset{\cdot}{h}(x, \ulcorner A\urcorner).$$

Here $\underset{\cdot}{h}$ represents a function satisfying the following equation:

$$h(n, \ulcorner B\urcorner) \;=\; \ulcorner\underbrace{\Box\ulcorner\Box\ulcorner\ldots}_{n}\ulcorner B\urcorner\underbrace{\urcorner\ldots\urcorner}_{n}\urcorner.$$

(See also Halbach [18, Lemma 4.1(iv)] for a proof of the example.) By a similar argument one can show that there is no PW-model based on $\langle\omega, <\rangle$.

Instead of McGee's theorem Visser's theorem [45] on illfounded hierarchies of languages can be used for deriving the above result. The relations between the different ω-inconsistency results and their implications are studied by Leitgeb [29].

Also Gödel's Second Incompleteness Theorem can be rephrased in terms of possible-worlds semantics. First we prove Löb's Theorem, which will be very useful in later sections.

Transitivity yields the predicate analogue of the K4 axiom:

LEMMA 10. *Let $\langle W, R, V\rangle$ be a PW-model based on a transitive frame. Then*

$$V(w) \models \Box\ulcorner A\urcorner \to \Box\ulcorner\Box\ulcorner A\urcorner\urcorner$$

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

LEMMA 11 (Löb's Theorem). *For every world w in a PW-model based on a transitive frame and every sentence $A \in \mathcal{L}_\Box$ the following holds:*

$$V(w) \models \Box\ulcorner\Box\ulcorner A\urcorner \to A\urcorner \to \Box\ulcorner A\urcorner.$$

PROOF. The proof is the usual proof of Löb's Theorem with the provability predicate replaced by the primitive symbol $\Box$. The diagonalisation lemma yields a sentence K:

$$V(w) \models K \leftrightarrow (\Box\ulcorner K\urcorner \to A). \tag{1}$$

Since this holds in any world, we may necessitate the sentence according to Lemma 3:

$$\begin{array}{ll} V(w) \models \Box\ulcorner K\urcorner \to (\Box\ulcorner\Box\ulcorner K\urcorner\urcorner \to \Box\ulcorner A\urcorner) & \\ V(w) \models \Box\ulcorner K\urcorner \to \Box\ulcorner A\urcorner & \text{Lemma 10} \\ V(w) \models (\Box\ulcorner A\urcorner \to A) \to (\Box\ulcorner K\urcorner \to A) & \\ V(w) \models (\Box\ulcorner A\urcorner \to A) \to K & \text{by (1)} \end{array}$$

Since this holds at all worlds w, we may necessitate the last sentence and obtain Löb's Theorem by applying $V(w) \models \Box\ulcorner K\urcorner \to \Box\ulcorner A\urcorner$. ⊣

Obviously Löb's Theorem is a kind of an induction principle (see Boolos [6]). Since we cannot assign arbitrary truth-values to the sentences of $\mathcal{L}_\Box$ in any given world, we cannot conclude that R is converse wellfounded in general.

We shall also use the following version of the theorem:

COROLLARY 12. *Assume $\langle W, R\rangle$ is transitive and $A \in \mathcal{L}_\Box$. If $V(w) \models \Box\ulcorner A\urcorner \to A$ for all $w \in W$, then also $V(w) \models A$ for all worlds w.*

From (the modal version of) Löb's Theorem 11 (the modal version of) Gödel's Second Incompleteness Theorem can be derived by setting $A = \bot$ for a fixed contradiction $\bot$. A dead end is a world that does not see any world. Obviously $V(w) \models \Box\ulcorner\bot\urcorner$ holds if and only if w is a dead end.

EXAMPLE 13. In a transitive frame admitting a valuation every world is either a dead end or it can see a dead end.

PROOF. Since the frame is transitive the predicate analogue of the K4 axiom scheme holds. This suffices for a proof of the formalised Incompleteness Theorem and we obtain $V(w) \models \Box\ulcorner\bot\urcorner \vee \Diamond\ulcorner\Box\ulcorner\bot\urcorner\urcorner$. ⊣

In the sequel we shall provide a theorem (Theorem 32) that implies all the above results and does away with the need for all the different diagonal sentences, which have to be found for every frame separately. In the end all limitative results can be derived from Löb's Theorem.

§4. The Characterisation Problem for Transitive Frames I. In the preceding section we have presented some examples of limitations on the existence of valuations. All frames discussed in the preceding section have infinite R-chains $w_1 R w_2 R w_3 R \dots$, that is, they are not converse wellfounded. It is not hard to show that all converse wellfounded frames support a PW-model (see Theorem 31 below). However, also some converse illfounded frames admit valuations. Because of these frames the Characterisation Problem is nontrivial.

The following two sections are devoted to a special case of the Characterisation Problem, namely to the following question: Which *transitive* frames admit valuations? In the present section we provide sufficient conditions for the existence of a valuation; some (unfortunately inequivalent) necessary conditions are then given in the next section. Throughout both sections we assume that R is transitive. We give here a brief outline of the results of these two sections and (some) of Section 6.

It is natural to look first at frames with an accessibility relation that is a converse wellfounded ordering, indeed a converse wellorder. For these it is unproblematic to assign valuations (see Theorem 31). From these one considers orderings that have some co-initial wellfounded part (Definition 14), and we seek to characterise (first) those linear orderings that have some converse wellordered part, with a possibly illfounded initial part. We first analyse however how valuations are built up along converse wellorderings. It turns out that there is a natural operator, Φ, associated to the transition of the extension of the $\Box$ symbol from one world to its immediate predessor. This operator is decreasing, and hence, if the converse wellordering is sufficiently long, results in a fixed point after some least ordinal number κ, of stages of application. The main theorem of the section (Theorem 23) is a calculation of κ. It is not hard to imagine, for the first stages at least, that the extension of $\Box$ along the co-initial segment of an accessibility relation, being a predicate referring to truths at previous worlds, in fact contains sufficient information to amount to an iteration of a truth predicate. Indeed this is precisely what happens; and once into the transfinite the idea is that the extensions of $\Box$ (called $X_{\alpha+1}$ at the "$\alpha+1$'st world") in fact code truth sets for the α'th level of the Gödel constructible hierarchy L_α. At a certain level of this hierarchy,

γ, the structure $\langle L_\gamma, \in \rangle$ is sufficiently closed that it becomes *first-order reflecting* (see the discussion before Theorem 23). It is precisely at this stage that we have our fixed point. The structure $\langle L_\gamma, \in \rangle$ is a model of *Kripke-Platek set theory*, KP. Such a theory is important for us since our (simplest to state) sufficient condition on those linear orders that carry valuations, are as those that are isomorphic to (initial segments of) the ordinals of models of KP. Such models need not be wellfounded. This is done at Theorem 27. A more technical sufficient condition is that of Theorem 25; a more quotable version is the Corollary 26: essentially if a KP model $\mathcal{A}$ contains a linear frame $\langle W, R \rangle$ as an element of the model, then that frame supports a valuation if $\mathcal{A}$ is a model of the statement that "R is converse wellfounded" (irrespective of the truth of that statement in the real word – as $\mathcal{A}$ may itself be illfounded, and hence assess the truth value of this assertion incorrectly).

We are unable to give an exact characterisation of all linearly ordered frames that admit valuations, however the main result of section 5 gives a necessary condition on what those linear orders must look like: if they are converse illfounded, then the converse wellfounded part must have order type greater than κ or be an *admissible ordinal* $\omega < \alpha \leq \kappa$ (Lemma 30). (That there are such frames is a standard fact and an application of the Barwise Compactness theorem - *cf* Lemma 35 and the related Proposition 37 of section 7 where it is shown that there are frames with converse wellfounded part any countable admissible ordinal $> \omega$, which admit a valuation containing a Φ-fixed point (necessarily by the above in the converse-illfounded part of the frame).)

It seems to us that the problem of finding which linear orders admit frames is a particular case of the problem "Along which linear orderings can one build L-hierarchies?", or equally, hierarchies of ramified analysis. We make one simple conjecture here concerning just recursive linear orderings (at the end of section 4). This amounts to the claim that the sufficient condition of Theorem 27 is also necessary; but more generally it seemed to us at one point reasonable as a conjecture that one could build such a frame $\langle W, R \rangle$ with an associated valuation by piecing together pseudowellorders, whose converse wellfounded parts were of increasing admissible order type, as one descended through R.

We now proceed to the formal definitions. As above, a world w is *converse wellfounded* in a frame $\langle W, R \rangle$ if and only if every subset of $w{\uparrow} =_{df} \{ u \in W \mid wRu \}$ has an R-maximal element. A set of worlds is converse wellfounded if and only if all its elements are converse wellfounded; thus a set S of worlds is converse wellfounded if and only if $\bigcup_{w \in S} w{\uparrow}$ is wellfounded with respect to R^{-1}.

Definition 14. For any partial order $\langle X, T \rangle$ we define the *wellfounded part* as $\mathrm{WFP}(X) = \mathrm{WFP}(\langle X, T \rangle) =_{df} \langle X', T{\restriction}X' {\times} X' \rangle$ where X' is the maximal set $X' \subseteq X$, downwards closed under T, on which T is wellfounded.

Associated with any wellorder is the natural *rank function* defined by transfinite recursion. We define the rank of the wellfounded part of any partial order to be the supremum of the ordinal ranks of its elements.

Concerning our frames, the *converse wellfounded part* of a set of the form $\{ v \mid wRv \}$ is then the largest R-upwards closed set X (i.e., the largest X such that $\forall y \in X(yRx \Rightarrow x \in X)$) on which R^{-1} is wellfounded. The *depth* of a world w is defined to be the rank of R^{-1} restricted to the converse wellfounded part of $\{ v \mid wRv \}$. Thus a world does not have to be converse wellfounded to have a depth. (Later we shall need to apply this definition also to frames that are not transitive. If R is not transitive, we set the depth of w in $\langle W, R\rangle$ as the depth of w in $\langle W, R^*\rangle$, where R^* is the transitive closure of R.) If R is transitive and converse wellfounded, then $V(w)$ is determined by the depth of w. Thus V may be defined inductively by an operator Φ, which we shall study now.

Φ will generate suitable extensions of $\Box$. Thus we start with the set $\mathcal{L}_\Box$ of all sentences (which, perforce, are all in the extension of $\Box$ at any dead end) and Φ will progressively remove sentences of $\mathcal{L}_\Box$ in order to define the extension of $\Box$ at worlds of larger depth. Thus the operator does not build up larger and larger sets from the empty set; it rather reduces $\mathcal{L}_\Box$. We could work instead with an operator generating suitable extensions for $\Diamond$ from the empty set, which we did in [19], but this makes the notation awkward.

Definition 15. We set for any set X of natural numbers:

$$\Phi(X) \ =_{df} \ X \cap \{ A \in \mathcal{L}_\Box \mid (\mathbb{N}, X) \models A \}.$$

In the following, $\Phi^\alpha(X)$ designates the α-fold application of Φ to X; at limit ordinals λ, $\Phi^\lambda(X)$ is defined as the intersection $\bigcap_{\alpha<\lambda} \Phi^\alpha(X)$ of all previous stages.

The following lemma is established by induction on the depth α of w (and is a special case of 31).

Lemma 16. *Assume $\langle W, R, V\rangle$ is a PW-model based on a transitive converse wellfounded frame. If w has depth α, then $V(w) = \Phi^\alpha(\mathcal{L}_\Box)$ holds for all ordinals α.*

A set $X \subseteq \mathcal{L}_\Box$ is a fixed point (of Φ) if and only if $\Phi(X) = X$. For any set $\Phi(X) \subseteq X$ holds and, consequently, $\Phi^\alpha(\mathcal{L}_\Box) \supseteq \Phi^\beta(\mathcal{L}_\Box)$ if $\alpha \leq \beta$. Thus by iterated applications of Φ to $\mathcal{L}_\Box$ we arrive at a fixed point:

Lemma 17. *There is an ordinal κ such that $\Phi^\kappa(\mathcal{L}_\Box) = \Phi^\alpha(\mathcal{L}_\Box)$ for all $\alpha \geq \kappa$.*

The operator Φ (as well as the corresponding operator for $\Diamond$) is not monotone, that is, $X \subseteq Y$ does not generally imply $\Phi(X) \subseteq \Phi(Y)$. Therefore the fixed point $\Phi^\kappa(\mathcal{L}_\Box)$ need not be maximal (and the fixed point generated by the corresponding operator for $\Diamond$ need not be minimal).

The operator Φ compares with the revision operator Ψ of revision semantics as introduced by Gupta [17] and Herzberger [21] (see Gupta and Belnap [5]). The revision operator Ψ takes a set X to the new extension $\{\, A \in \mathcal{L}_\Box \mid (\mathbb{N}, X) \models A \,\}$. Since $X \subseteq \Psi(X)$ does not obtain, at a limit level one cannot simply take the intersection of all levels up to the limit level. Several alternative treatments of limit stages have been proposed.

If we start with the set $\mathcal{L}_\Box$ of all sentences, apply Φ repeatedly to it and take intersections at limit stages, then the liar sentence L with $L \leftrightarrow \neg\Box\ulcorner L \urcorner$ is in $\mathcal{L}_\Box$ but not in any $\Phi^\alpha(\mathcal{L}_\Box)$ for $\alpha \geq 0$. In revision semantics, in contrast, the liar sentence will always flip in and out of the extension of the truth predicate, thereby causing trouble with the definition of the extension at limit stages; in particular, as already noted, the revision process will never converge.

Fixed points of Φ can be characterised in the following way.

Lemma 18. *$X \subseteq \mathcal{L}_\Box$ is a fixed point if and only if $X \models \Box\ulcorner A \urcorner \to A$ for all $A \in \mathcal{L}_\Box$.*

The existence of fixed points implies that there are frames that are not converse wellfounded but which admit a valuation nevertheless.

Theorem 19. *Assume $\langle W, R\rangle$ is transitive and every converse illfounded world in $\langle W, R\rangle$ has depth at least κ. Then $\langle W, R\rangle$ admits a valuation.*

Proof. The definition of V for converse wellfounded w is forced by Lemma 16. If $w \in W$ has depth at least κ we put $V(w) = \Phi^\kappa(\mathcal{L}_\Box)$. This holds in particular for all converse illfounded worlds.

By Lemma 17 the condition

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

is satisfied by all worlds $w \in W$. ⊣

In the following, ω_1 will always denote the first nonrecursive ordinal (see Rogers [39]). Using an ordinal notation system for the ordinals below ω_1 it is possible to construct sentences that hold at a world w of a transitive frame if and only if w is converse wellfounded and has depth $\alpha < \omega_1$. However, in general one cannot express converse wellfoundedness in $\mathcal{L}_\Box$.

Corollary 20. *If $\langle W, R, V\rangle$ is a transitive converse illfounded model, then there is no formula A that holds exactly in the converse wellfounded worlds of $\langle W, R\rangle$.*

Proof. According to Theorem 19 there are frames $\langle W, R\rangle$ with not converse wellfounded worlds that support a PW-model. If there were such a formula A,

$$V(w) \models \Box\ulcorner\Box\ulcorner A\urcorner \to A\urcorner$$

would hold in every world $w \in W$. By Lemma 11 (Löb's Theorem) $\Box\ulcorner A\urcorner$ would hold in every world of W. But by assumption A does not hold in these converse illfounded worlds. Contradiction! ⊣

We determine the exact size of the closure ordinal κ of Φ and thereby show that there is a gap between ω_1 and κ.

We abbreviate by setting $X_\alpha = \Phi^\alpha(\mathcal{L}_\Box)$. The function $\delta \mapsto X_\delta$ can be defined by a Σ_1-recursion over the ordinals of $\mathcal{M}$ where the latter is any sufficiently closed model $\langle \mathcal{M}, E\rangle$ containing the standard model $\mathbb{N}$ of arithmetic. In particular this is true over any model $\mathcal{M}$ of Kripke–Platek (KP) set theory containing the integers as a standard set. For background on the theory KP and on admissible sets, see Barwise [3]. We review some of this theory and these notions here. We shall assume the reader is familiar with Gödel's constructible hierarchy, which can also be found in [3]. The theory of KP is expressed in the language of set theory, that is, with equality, and the single nonlogical symbol $\dot\in$.

Transitive models $\mathcal{M} = \langle \mathcal{M}, \in\rangle$ of KP, with the symbol $\dot\in$ getting its standard interpretation, are called *admissible sets.* We shall have occasion to consider in the sequel nonwellfounded models of KP: that is, models of the form $\langle \mathcal{A}, E\rangle$ where the epsilon relation E of $\mathcal{A}$ is not wellfounded. The *wellfounded part* of such a model $\mathrm{WFP}(\mathcal{A}) =_{df} \mathrm{WFP}(\mathcal{A}, E)$ is defined in accordance with Definition 14, by considering E as a partial ordering on the domain of $\mathcal{A}$.

If α is the ordinal rank of the wellfounded part $\mathrm{WFP}(\mathcal{A})$ (we shall write $\alpha = \rho(\mathrm{WFP}(\mathcal{A}))$), then it is known that $\mathrm{WFP}(\mathcal{A})$ is a model of KP (the so-called "Truncation Lemma", see [3, II 8.4]). In particular this lemma implies that the constructible part of $\mathrm{WFP}(\mathcal{A})$, which is isomorphic to $\langle L_\alpha, \in\rangle$, also is a model of KP. If $\mathcal{A}$ is a model of KP, $On(\mathcal{A})$ denotes the class of ordinals in the sense of $\mathcal{A}$. We adopt the convention that we identify the wellfounded part of $On(\mathcal{A})$ with the standard ordinals to which they are order-isomorphic (and similarly we identify the E-relation restricted to the wellfounded part as the actual $\in$-relation). Hence $On(\mathrm{WFP}(\mathcal{A})) \supseteq \omega+1$ holds if and only if the integers of $\mathcal{A}$ are standard. We shall only consider in what follows models of KP in which the integers are standard and moreover form a set.

The feature, and strength, of KP which we shall be exploiting is its ability to define functions by transfinite recursion along relations (usually the ordinals of the model, or its E-relation), if the defining clauses of the recursion are sufficiently simple: that is, have a form which is—provably in KP—equivalent to a Σ_1-form. Thus ordinal arithmetic, and in particular the Gödel L-hierarchy, can be defined by suitable Σ_1-recursions over any model $\mathcal{A}$ of KP. Hence such models $\mathcal{A}$ have versions of the L-hierarchy contained inside them as "inner models" of KP. Thus if the model $\mathcal{A}$ is illfounded, the L-hierarchy of $\mathcal{A}$ coincides with L_α on the wellfounded part of $\mathcal{A}$ (by our convention) but there will also be sets $Z \in \mathcal{A}$ that are the levels L_a of the constructible hierarchy of $\mathcal{A}$ where a is a nonstandard ordinal in the illfounded part of $\mathcal{A}$. Another example is the operation of Φ: this can be defined by a Σ_1-recursion inside any KP-model $\mathcal{A}$ using as frame the converse of the $\mathcal{A}$-ordinals, $\langle On(\mathcal{A}), E^{-1}\restriction On(\mathcal{A})^2\rangle$. The latter may be illfounded, and if so, the result will be an illfounded frame with a valuation defined in $\mathcal{A}$. (It is to be emphasised that $\mathcal{A}$ of course thinks its ordinals are wellfounded: only in the "real world" of all sets do we see that there is a counterexample to wellfoundedness: that is, a nonempty set $X \subseteq On(\mathcal{A})$ which has no E-minimal element. It is a consequence of KP that there is no such $X \in \mathcal{A}$.)

We shall always assume that the Axiom of Infinity is part of KP. An ordinal is *admissible* if $\langle L_\alpha, \in\rangle$ is a model of KP. For $R \in 2^\omega$ (or equivalently $R \subseteq \omega$) we let ω_1^R denote the supremum of the ranks of all wellorderings of ω that are recursive in R. ADM is the class of all admissible ordinals, which by this assumption are all greater than ω. ADM is not a closed class; we thus define ADM* as ADM together with all of its limit points.

The idea of the proof of the following theorem is that below the closure ordinal κ of the operation Φ, the sets $X_{\delta+1}$ (for $\delta \geq \omega$) can be construed essentially as truth sets for the δ'th level of the L-hierarchy. The point where this identification breaks down is precisely where the L-hierarchy requires parameters in the first-order definitions that are required to make up the elements of the next level.

Let γ be least so that L_γ has a transitive Σ_1-end extension (that is, there is a transitive M with L_γ as an element, and $\langle L_\gamma, \in\rangle \prec_{\Sigma_1} \langle M, \in\rangle$). This ordinal is in some senses a large one: it is recursively Mahlo (indeed a fixed point in a recursively "hypermahlo" hierarchy). It is easy to see that $L_\gamma \models \forall x(|x| = \omega)$. Further one may show the following:

Proposition 21. *γ is the least ordinal for which $\langle L_\gamma, \in\rangle \prec_{\Sigma_1} \langle L_{\gamma+1}, \in\rangle$ holds.*

Actually we could have used this as our definition of γ in the sequel. From this it follows that γ is less than the least ordinal ν so that $L_\nu \models$ KP + Σ_1-separation.

Lemma 22. *L_γ is an admissible set that is the Skolem Hull of Σ_1-parameter-free terms inside itself. Thus any element x of L_γ is named by a parameter-free Σ_1 Skolem term, and the $\in$-diagram of L_γ is essentially given by the Σ_1-truth set for L_γ.*

This latter property fails first at $L_{\gamma+1}$. We sketch a proof of the lemma.

PROOF. That $L_\gamma \models \mathrm{KP}$ is fairly immediate. (If, for example, some instance of Δ_0-comprehension failed, one obtains a function $g\colon x \to On$, for some $x \in L_\gamma$, with g Σ_1-definable over L_γ, that is, unbounded in γ. However $L_{\gamma+1}$ believes $\mathrm{ran}(g)$ *is* bounded, and this Σ_1-statement transfers down to L_γ—a contradiction.) If the Skolem Hull property failed for some set $x \in L_\gamma$ we may use a function F that is $\Delta_1^{L_\gamma}$-definable and is a bijection between γ and L_γ, to assume, without loss of generality, that some least ordinal $\delta < \gamma$ fails to be Σ_1-definable. But then no ordinal η with $\delta < \eta < \gamma$ can be Σ_1-definable over L_γ without parameters. If η were a counterexample, then we use the fact that in $L_{\eta+1}$ there is a least (in the canonical wellorder of $L_{\eta+1}$) wellorder of ω, S say, in order type η. There must be such a wellorder: otherwise η is a cardinal inside $L_{\eta+1}$ and then it can be shown there are chains of fully elementary chains of models of ZF^--models of the form $L_{\gamma_n} \prec L_{\gamma_{n+1}}$ with all $\gamma_n < \eta < \gamma$. Thus δ would be Σ_1-definable from this wellordering S (as being isomorphic to the "rank of k in S" for some $k \in \omega$) in L_γ. Putting this together with the Σ_1-definition of η, and the further Σ_1-definition of $L_{\eta+1}$ from η, would constitute a parameter-free Σ_1-definition of δ—a contradiction.

Now it can easily be shown that $L_\delta \prec_{\Sigma_1} L_\gamma$. If this failed, then for some Σ_0-formula $\psi(v_0, v_1)$ we should have $L_\gamma \models \exists x\, \psi[x, y]$ but $L_\delta \models \exists x\, \psi[x, y]$ fails for some set $y \in L_\delta$. But y has a Σ_1-parameter-free definition in L_γ, so in effect we can drop y by rewriting the definition $\exists v_0\, \psi(v_0, v_1)$ as $\exists v_0\, \psi'(v_0)$. Now γ is a limit ordinal, hence the "least ordinal η so that $L_\eta \models \exists x\, \psi'[x]$" is a Σ_1-definition of an ordinal greater than δ. Contradiction!

But γ was least so that L_γ had a Σ_1-end extension! Hence there is no such δ. ⊣

The two facts from Proposition 21 and Lemma 22 above from the folklore are used in the proof of Theorem 23 below.

It is also well known that γ is least so that L_γ is *first-order reflecting*, that is, for all n, if φ is any Π_n-formula (with parameters from L_γ allowed) in the language of set theory, then the following holds (see Aczel [1]):

$$L_\gamma \models \varphi \implies \exists \alpha < \gamma \;\; L_\alpha \models \varphi.$$

(As indicated above, L_γ is a model of a strong extension of KP—stronger than the theory KPi which is proof-theoretically equivalent to ACA + Bar Induction + Δ^1_2-Comprehension Scheme; however, it is known, by consideration of (i) above, that it is weaker than the latter strengthened with the Π^1_2-Comprehension Scheme—see Rathjen [37].)

THEOREM 23. $\gamma = \kappa$.[1]

In particular, this shows that $\kappa > \omega_1$ (like any admissible ordinal, ω_1 is Π_2-reflecting but it is not Π_3-reflecting). We shall prove this theorem after a preliminary lemma.

We shall let Σ_n-$\mathrm{Th}(L_\beta)$ denote the set of Σ_n-sentences true in $\langle L_\beta, \in\rangle$. We use lower-case Greek letters φ, ψ etc. for formulae in the set-theoretic language possibly expanded by additional parameters. Note if Σ_1-$\mathrm{Th}(L_\delta)$ is Σ_k-definable over $(\mathbb{N}, X_\delta)$, say,

$$\varphi \in \Sigma_1\text{-}\mathrm{Th}(L_\delta) \iff (\mathbb{N}, X_\delta) \models C_1(\ulcorner\varphi\urcorner), \tag{2}$$

then it is well known that Σ_n-$\mathrm{Th}(L_\delta)$ is Σ_{k+n}-definable over $(\mathbb{N}, X_\delta)$ in a definition effectively obtainable from that of C_1. We shall see by the following Lemma that we may take here $k = 1$ and thus we shall have for a suitable formula $C_n(x)$:

$$\varphi \in \Sigma_n\text{-}\mathrm{Th}(L_\delta) \iff (\mathbb{N}, X_\delta) \models C_n(\ulcorner\varphi\urcorner). \tag{3}$$

LEMMA 24. *Σ_1-Th(L_δ) is Σ_1-definable over $\langle\mathbb{N}, X_\delta\rangle$ (and thus is recursively enumerable in X_δ). The definition is uniform for all $\delta+1$, that is, it is independent of δ. It is similarly uniform for all limits δ.*

PROOF. We prove the lemma by induction on δ.

If δ is a limit ordinal, we have the following by the upwards persistence of Σ_1-sentences:

$$\Sigma_1\text{-}\mathrm{Th}(L_\delta) \;=\; \bigcup_{\alpha<\delta} \Sigma_1\text{-}\mathrm{Th}(L_\alpha).$$

Thus we arrive at the following:

$$\begin{aligned} \varphi \in \Sigma_1\text{-}\mathrm{Th}(L_\delta) &\iff \exists\alpha < \delta \;\; \varphi \in \Sigma_1\text{-}\mathrm{Th}(L_\alpha) \\ &\iff \exists\alpha < \delta \;\; (\mathbb{N}, X_\alpha) \models C_1(\ulcorner\varphi\urcorner) \qquad\qquad \text{by (2)} \\ &\iff \exists\alpha < \delta \;\; (C_1(\ulcorner\varphi\urcorner)) \in X_{\alpha+1} \\ &\iff (C_1(\ulcorner\varphi\urcorner)) \in X_\delta. \end{aligned}$$

Therefore Σ_1-$\mathrm{Th}(L_\delta)$ is Σ_1 over $(\mathbb{N}, X_\delta)$.

We turn to the successor case. Assume φ is Σ_1 and $L_{\delta+1} \models \varphi$. Suppose $\varphi \equiv \exists y\, \psi(y)$ where $\psi(v_0)$ is Σ_0 with just v_0 free. Let $y \in L_{\delta+1}$ be a witness to this. Hence $L_{\delta+1} \models \psi(\overline{y})$. By construction $y \in \mathrm{Def}(\langle L_\delta, \in\rangle)$ and has therefore a first-order definition:

$$y \;=\; \{\, z \in L_\delta \mid L_\delta \models \chi(\overline{z}) \,\}.$$

By our assumption on γ (Fact (ii) above), we can assume all parameters that would normally be required in such a defining formula $\chi(v)$ have in

[1]The closure ordinal γ for the operator Φ was independently suggested to Volker Halbach by Peter Aczel, and by Philip Welch. [V.H.]

fact been replaced by Σ_1-defining terms. We may thus assume $\chi(v)$ has just the one free variable shown, and no other parameters from L_δ. Let $\psi(\chi)$ be the Σ_m-formula (for some m) obtained by replacing each mention of v in $\psi(v)$ by $\{\, u \mid \chi(u) \,\}$. For any formula ζ of set theory set $lh(\zeta)$ to be the smallest n such that ζ is in Σ_n. We thus have

$$\begin{aligned} \varphi \in \Sigma_1\text{-Th}(L_{\delta+1}) &\iff L_{\delta+1} \models \exists y\, \psi \\ &\iff \exists \chi \left(\psi(\chi) \in \Sigma_{lh(\psi(\chi))}\text{-Th}(L_\delta) \right) \\ &\iff \exists \chi \left(C_{lh(\psi(\chi))}(\ulcorner \psi(\chi) \urcorner) \in X_{\delta+1} \right). \end{aligned}$$

The last line follows from (3) and our inductive hypothesis. ⊣

PROOF OF THEOREM 23. We prove $\kappa \leq \gamma$ first. Suppose $\gamma < \kappa$ for a contradiction. Then $X_\gamma \neq X_{\gamma+1}$. Let $A \in X_{\gamma+1} \setminus X_\gamma$. Suppose that $A \in \Sigma_k$. Then recall that

$$\left\{ A \;\middle|\; A \in \Sigma_k \wedge (\mathbb{N}, X_\gamma) \models A \right\}$$

is recursive in ${X'_\gamma}^{(k)}$ (the k'th Turing jump of X_γ). By the preliminary remarks above, as L_γ is a KP-model, $X_\gamma = \bigcup_{\alpha<\gamma} X_\alpha$ is $\Sigma_1^{L_\gamma}$-definable. Then ${X'_\gamma}^{(k)}$ is $\Sigma_{k+1}^{L_\gamma}$-definable, and is in $L_{\gamma+1}$. Thus γ is Σ_1-definable in $L_{\gamma+1}$ as "the least $\bar{\gamma}$ so that $n \in {X'_{\bar{\gamma}}}^{(k)}$". But then L_γ cannot be a Σ_1-substructure of $L_{\gamma+1}$. Thus we have $\kappa \leq \gamma$.

It remains to show $\kappa \geq \gamma$. We derive a contradiction from $\kappa < \gamma$ by assuming the following:

$$\varphi \;\in\; \Sigma_1\text{-Th}(L_{\kappa+1}) \setminus \Sigma_1\text{-Th}(L_\kappa).$$

Such a φ exists, since otherwise $L_\kappa \prec_{\Sigma_1} L_{\kappa+1}$, whereas γ is the least ordinal with this property. Thus by Lemma 24 we arrive at the following:

$$\exists \chi \left(C_{lh(\psi(\chi))}(\ulcorner \psi(\chi) \urcorner) \in X_{\kappa+1} \setminus X_\kappa \right),$$

whereas κ is the closure ordinal for this operation and X_κ is a fixed point. Contradiction! ⊣

Thus the characterisation of the class of transitive frames admitting valuations yielded by Lemma 32 and Theorem 19 is incomplete because it does not say whether worlds of depth α with $\omega_1 \leq \alpha \leq \kappa$ have to be converse wellfounded. We shall now show that these worlds may be illfounded, too.

We note that if we concentrate on countable frames $\langle W, R \rangle$, then we may without loss of generality assume that R is a binary relation on $W = \omega$. If R is such a binary relation, its characteristic function $r \in 2^\omega$ is defined as follows:

$$\text{for all } n, m\colon \quad r(\langle n, m \rangle) = 1 \;\leftrightarrow\; nRm.$$

We add a symbol $\overline{r}$ to the language of KP and numerals $\overline{n}$ for all natural numbers. Diag(R) contains the sentence $\overline{r} \in 2^{\omega}$ as well as all sentences $\overline{r}(\langle\overline{n},\overline{m}\rangle) = \overline{k}$ such that $r(\langle n,m\rangle) = k$. For $\mathcal{A}$ a KP-model we let $\kappa_{\mathcal{A}} \simeq$ the closure ordinal of the operation Φ—as defined inside $\mathcal{A}$—if it exists; otherwise it is undefined.

THEOREM 25. *Assume $\langle\omega, R\rangle$ is transitive and the theory* KP+Diag(R) *does not prove in ω-logic the following conjunction:*

(i) "$\overline{r}$ codes a converse illfounded relation $\bar{R}$ on ω," and:
(ii) "if $\bar{\kappa}$, the closure ordinal of Φ, exists, then there exists a converse illfounded world w with depth$(w) < \bar{\kappa}$."

Then $\langle\omega, R\rangle$ admits a valuation V and thus is a frame with a PW-model.

PROOF. As the quoted assertion is not provable, let $\mathcal{A}$ be a KP-model containing r, in which the integers are standard, and in which the quoted assertion is false. Note that $\bar{\kappa}_{\mathcal{A}}$, if it exists, could be a nonstandard ordinal of $\mathcal{A}$.

It suffices to note that $\mathcal{A} \models$ "$Z = \Phi(X)$" if and only if $Z = \Phi(X)$. Thus in $\mathcal{A}$ the Σ_1-function $F(w) = X_w$ is defined along the part of the relation $\bar{R}^{-1}$ that $\mathcal{A}$ believes to be wellfounded. The function so defined in $\mathcal{A}$ satisfies these defining clauses for $w \mapsto V(w)$ in the real world. If $\kappa_{\mathcal{A}} = \bar{\kappa}$ exists, as long as $\mathcal{A}$ believes the converse wellfounded part has length $\bar{\kappa}$, one may assign $V(w) = \Phi^{\bar{\kappa}}(\mathcal{L}_{\Box})$ for those w in R^{-1} of depth $\geq \bar{\kappa}$. $\dashv$

We obtain the following corollary from the proof of the last theorem. (Compare this with Theorem 19):

COROLLARY 26. *Let $\langle W, R\rangle$ be transitive, $\mathcal{A}$ be a model of* KP, *$R \in \mathcal{A}$ and suppose if $\mathcal{A} \models$ "$\langle W, R\rangle$ has a converse illfounded world w", then $\mathcal{A} \models$ "$\kappa_{\mathcal{A}}$ exists, and depth$(w) \geq \kappa_{\mathcal{A}}$". Then $\langle W, R\rangle$ admits a valuation.*

We can add a somewhat more perspicuous example in the case of linear orderings.

THEOREM 27. *Let R be a linear ordering of ω. Let $\langle W, R\rangle$ be the frame with $W = \omega$. Then*

$\langle W, R\rangle$ admits a valuation if $\langle W, R^{-1}\rangle$ is a pseudo-wellordering.

We make some remarks on the notions involved. $\langle\omega, S\rangle$ is a *pseudo-wellorder* if (a) it is a linear ordering and (b) if $X \subseteq \omega$ is an S-descending chain, then X is not hyperarithmetic in S. Thus, in particular, a recursive pseudo-wellorder can have no hyperarithmetic descending chains; pseudo-wellorders for recursive S were introduced by Spector and Feferman [12] and studied by J. Harrison who showed in [20] the following:

THEOREM 28 (Harrison). *Let η be the order type of the rationals. Let $\langle\omega, S\rangle$ be a linear ordering. This ordering is a pseudo-wellordering if and only if it has order type $\omega_1^S(1+\eta) + \alpha$ where $\alpha \leq \omega_1^S$.*

(In [20] again, only the special case of S a recursive linear ordering is studied, but the argument straightforwardly relativises.) Pseudo-wellorders characterise the order types of countable models of KP. One may show the following fact: $\langle \omega, S\rangle$ is a pseudo-wellorder if and only if there is a countable model $\mathcal{A}$ of KP with $\langle \omega, S\rangle$ order-isomorphic to an initial segment of $On(\mathcal{A})$. The *character* of a pseudo-wellorder is essentially the admissible ordinal ω_1^S.

Proof of Theorem 27. Let $\langle \omega, R\rangle$ be a frame with R^{-1} a pseudo-wellordering. Let $\mathcal{B}$ be a KP-model with $\langle \omega, R^{-1}\rangle$ isomorphic to an initial segment of $On(\mathcal{B})$. (If R^{-1} is not a wellorder, we construct such a $\mathcal{B}$ below at Proposition 37 with $\mathrm{WFP}(On(\mathcal{B}))$ isomorphic to ω_1^R.) But now arguing as in Theorem 25 we can establish a valuation function $F(w) = X_w$ on this initial segment. This suffices. $\dashv$

Conjecture. Let R be a recursive linear ordering of ω. Then $\langle \omega, R\rangle$ is a frame admitting a valuation if and only if R^{-1} is a pseudo-wellorder.

Note that in this case, as R is recursive, $\omega_1^R = \omega_1$. We could have made a more general conjecture about the totality of countable pseudo-wellorders: it seems to us that, ignoring the final short ordinal part of a pseudo-wellordering, it should be possible to assemble a converse linear ordering admitting a valuation that is built from pseudo-wellorders of increasing character. However we rest content with the above conjecture.

§5. The Characterisation Problem for Transitive Frames II. We now turn to necessary conditions for the existence of a valuation.

As above, ADM* is the class ADM of all admissible ordinals (without ω) together with its limit points.

Theorem 29. *Let $\langle W, R, V\rangle$ be a transitive model. Then the following must hold for the depth α of every converse illfounded world in $\langle W, R\rangle$: either $\alpha \in \mathrm{ADM}^*$ or $\alpha \geq \kappa$.*

This follows by general considerations, and a similar argument leads to the following lemma, that proves essentially this result restricted to linear orders.

Lemma 30. *Suppose $\langle W, R\rangle$ is a frame with R^{-1} an illfounded linear ordering which admits a valuation V. Then R^{-1} has an initial wellordered segment of order type $\alpha \in \mathrm{ADM}$ or $\alpha \geq \kappa$.*

Proof. By Example 13, R^{-1} has a nonempty wellordered initial segment. The maximal wellordered initial segment of R^{-1} is isomorphic to an ordinal $\langle \alpha, <\rangle$. We identify this initial wellordered part with α itself. We suppose from now on that R^{-1} has an illfounded part beyond this segment and that $\alpha < \kappa$. Thus we need to prove $\alpha \in \mathrm{ADM}$.

First we show that α cannot be a successor ordinal $\beta+1$. In order to arrive at a contradiction, assume that β is the maximal element of the wellordered initial segment. Since $\alpha < \kappa$, there is a sentence $B \in \mathcal{L}_\Box$ satisfying the following condition:

$$V(\beta) \models \Box\ulcorner B\urcorner \wedge \neg B. \tag{4}$$

For $\delta \leq \beta$ we have:

$$V(\delta) \models \Box\ulcorner B\urcorner. \tag{5}$$

Assume $w \in W$ is converse illfounded. Since there are infinitely many worlds between w and the wellordered initial segment, we conclude from (4) the following:

$$V(w) \models \neg\Box\ulcorner\Box\ulcorner B\urcorner\urcorner. \tag{6}$$

Together, (5) and (6) yield the following for every world $w \in W$:

$$V(w) \models \Box\ulcorner\Box\ulcorner B\urcorner\urcorner \to \Box\ulcorner B\urcorner.$$

Löb's Theorem (Corollary 12) yields $V(w) \models \Box\ulcorner B\urcorner$. But this is a contradiction because $\Box\ulcorner B\urcorner$ fails at all converse illfounded worlds.

Therefore α has to be a limit number and, by assumption, α is inadmissible over $\langle L_\alpha, \in\rangle$. Hence by Theorem 23 for the closure ordinal κ there is a function that has domain ω, is cofinal in the ordinal α, and is Σ_1-definable over L_α. Let us suppose it is defined by the formula $\psi(n,\gamma)$. As $\alpha < \kappa$ we need no parameters in this definition. We employ the formula $C_1(x)$ used in the proof of Lemma 24 and define the sentence D as

$$\exists n\, \neg C_1(\ulcorner \exists y\, \psi(\dot{n}, y)\urcorner).$$

As in the successor case, we shall show for the sentence D defined in this way that the following holds at all worlds $w \in W$:

$$V(w) \models\ \Box\ulcorner\Box\ulcorner D\urcorner\urcorner \to \Box\ulcorner D\urcorner. \tag{7}$$

In order to show

$$\forall\beta < \alpha \quad V(\beta) \models \Box\ulcorner D\urcorner, \tag{8}$$

consider any $\beta < \alpha$; as ψ does not define a total function over L_β, we have

$$\exists n \quad \big(\exists y\, \psi(\overline{n}, y)\big) \notin \Sigma_1\text{-Th}(L_\beta).$$

But, using (2) in the proof of Theorem 23, this is equivalent to

$$(\mathbb{N}, X_\beta) \not\models \forall n\, C_1(\ulcorner \exists y\, \psi(\dot{n}, y)\urcorner)$$

and (8) follows. (7) will be established, if we have shown that for all converse illfounded worlds $w \in W$ the following holds:

$$V(w) \models \neg\Box\ulcorner\Box\ulcorner D\urcorner\urcorner. \tag{9}$$

This may be seen as follows:

$$\begin{array}{ll}
\forall n \in \omega \;\; \big(\exists y\, \psi(\overline{n}, y)\big) \in \Sigma_1\text{-Th}(L_\alpha) & \text{definition of } \psi(n,\gamma) \\
\forall n \in \omega \;\; (\mathbb{N}, X_\alpha) \models C_1(\ulcorner \exists y\, \psi(\overline{n}, y) \urcorner) & \text{by (2)} \\
(\mathbb{N}, X_\alpha) \models \forall n\, C_1(\ulcorner \exists y\, \psi(\dot{n}, y) \urcorner) & \\
(\mathbb{N}, X_\alpha) \models \neg D. &
\end{array}$$

Since w can see all converse wellfounded worlds $\beta < \alpha$, we have $V(w) \subseteq X_\alpha = \bigcap_{\beta<\alpha} X_\beta$ and also (9) obtains because there are infinitely many (and thus at least two) worlds between w and the wellfounded part of R^{-1}. ⊣

The partial characterisation of the class of transitive frames admitting valuations yielded by Lemma 29 and Theorem 19 is incomplete because it does not say whether worlds of depth α with $\omega_1 \leq \alpha \leq \kappa$ have to be converse wellfounded. Proposition 37 below will show that for any admissible ordinal $\alpha > \omega$, there exist frames with converse illfounded worlds of depth α.

To summarise: For transitive frames $\langle W, R\rangle$: (i) If every world $w \in W$ has a depth $\alpha < \omega_1$ then R is converse wellfounded iff $\langle W, R\rangle$ admits a valuation; (ii) if every converse illfounded world has depth $\alpha \geq \kappa$ (without any restriction being placed on the ordering R for the part containing those worlds of depth $\geq \kappa$) then the frame admits a valuation. Thus for frames where all depths are smaller than ω_1 or larger than κ the Characterisation Problem for transitive frames is settled. Between these two ordinals the depth of every converse illfounded world of a transitive frame admitting a model must be admissible or a limit of admissibles according to Theorem 29. The problem of finding which, say, linearly ordered frames admit models appears to be part of a more general underlying question: which linear orderings support "constructible" hierarchies?

§6. The General Characterisation Problem. We now turn to the general Characterisation Problem, that is, we drop the assumption that the frames are transitive.

If $\langle W, R\rangle$ is converse wellfounded, then one can easily define a valuation for $\langle W, R\rangle$ by induction along R in the following way:

$$V(w) \;=\; \big\{\, A \in \mathcal{L}_\Box \;\big|\; \forall v (wRv \;\Rightarrow\; V(v) \models A) \,\big\}. \tag{10}$$

Obviously there is no choice in the definition of V. This implies the following sufficient condition for the existence of a valuation, which was proved by Gupta and Belnap in [5, Theorem 6E.5] in a slightly different form.

Theorem 31 (Gupta & Belnap). *Every converse wellfounded frame admits exactly one valuation.*

The necessary condition Theorem 29 for the existence of a valuation can be generalised in the following way:

THEOREM 32. *If $\langle W, R\rangle$ supports a PW-model, then in the transitive closure $\langle W, R^*\rangle$ of $\langle W, R\rangle$ every not converse wellfounded world in W has depth α with $\alpha \in \mathrm{ADM}^*$ or $\alpha \geq \kappa$.*

PROOF. There exists a two-place recursive function f that yields applied to a number n and a sentence A the sentence A itself preceded by n necessity predicates, i.e.:

$$f(y, \ulcorner A \urcorner) \;=\; \underbrace{\Box \ulcorner \Box \ulcorner \ldots \Box}_{y \text{ times}} \ulcorner A \urcorner \ldots \urcorner.$$

Instead of $\Box f(y, \ulcorner A \urcorner)$ we write $\Box^{y} \ulcorner A \urcorner$. Thus $\Box^{0} \ulcorner A \urcorner$ is $\Box \ulcorner A \urcorner$.

We define a new modal predicate corresponding to the transitive closure of that accessibility relation:

$$\Box^* x \;:\Longleftrightarrow\; \forall n\, \Box^n x. \tag{11}$$

Clearly, $V(w) \models \Box^{*} \ulcorner A \urcorner$ holds if and only if $V(v) \models A$ for all v with wR^*v, where R^* is the transitive closure of R.

Applying Theorem 29 to $\Box^*$ yields the claim. $\dashv$

Theorem 32 can be used for demonstrating that the frame $\langle \omega, Suc\rangle$ in Example 9 does not admit a valuation. Thus Theorem 32 implies all negative results in section 3. Therefore, in a sense, Löb's Theorem implies the possible-worlds analogues of Tarski's, Montague's, McGee's and Visser's theorems. In section 8 we shall show how to get from these analogues back to ω-inconsistency results.

By Theorem 19 a transitive frame admits a valuation if every converse illfounded world $w \in W$ has at least depth κ. One might hope to generalise Theorem 19 by showing that a frame admits a valuation if all not converse wellfounded worlds are of depth at least κ in the transitive closure of the frame. However, this generalisation fails. Thus there is no characterisation of the class of frames that admit valuations in terms of the transitive closure of the frames. This is shown in the next theorem.

THEOREM 33. *There are frames that do not support a PW-model, although their transitive closure does.*

PROOF. We add to the frame $(\kappa + 1, <)$ a world w that sees only the fixed-point world κ and itself, but not any worlds $\alpha < \kappa$. We call this frame $\langle W, R\rangle$.

The frame is not transitive because w cannot see the worlds seen by κ.

By Theorem 19 the transitive closure of the extended frame admits a valuation because the only converse illfounded world w of $\langle W, R^*\rangle$ has depth $\kappa + 1$.

D is the usual liar sentence with

$$D \leftrightarrow \neg\Box\ulcorner D\urcorner.$$

We shall show that the assumption that this frame admits a valuation leads to a contradiction.

Since κ sees a dead end, D holds at κ:

$$V(\kappa) \models D. \tag{12}$$

We show that both $V(w) \models D$ and $V(w) \models \neg D$ lead to a contradiction.

1. *Case*: $V(w) \models D$. Then we have also $V(w) \models \neg\Box\ulcorner D\urcorner$ and $\neg D$ must hold either at κ or at w, because w does not see any further world. From (12) we conclude $V(w) \models \neg D$, which contradicts the assumption.

2. *Case*: $V(w) \models \neg D$. From $V(w) \models \Box\ulcorner D\urcorner$ and the supposition that wRw we derive a contradiction again. ⊣

Obviously, if $\langle W, R, V\rangle$ is a PW-model and wRw, then $V(w)$ is a fixed point of Φ. So far we have used only the fixed point $\Phi^\kappa(\mathcal{L}_\Box)$. However, other fixed points of Φ are required for valuations of certain frames that are not transitive. We give an example.

Proposition 34. *There is a model with two reflexive worlds v and w such that $V(v) \neq V(w)$.*

Proof. The set of worlds is defined as follows:

$$W = \{\, \langle \alpha, k\rangle \mid \alpha \leq \kappa \text{ and } k \in \{0,1\}\,\}.$$

A world $\langle \alpha_1, k_1\rangle$ can see a world $\langle \alpha_2, k_2\rangle$ iff one of the following conditions is satisfied:

(i) $k_1 = k_2$ and $\alpha_2 < \alpha_1$;
(ii) $\alpha_1 = 0$, $k_1 = 1$, $\alpha_2 = \kappa$ and $k_1 = 0$;
(iii) $\alpha_1 = \kappa$ and $\alpha_2 = \kappa$ and $k_1 = k_2$.

Both copies $\langle \kappa, 0\rangle$ and $\langle \kappa, 1\rangle$ of κ can see themselves.

A valuation V is easily construed: it is the same as the valuation for the converse wellfounded frame (where $\langle \kappa, 0\rangle$ and $\langle \kappa, 1\rangle$ cannot see themselves) that is obtained by dropping clause (iii) above. For the κ of the first copy we have $V(\langle \kappa, 0\rangle) \models \Diamond\ulcorner 0 \neq 0\urcorner$ while for the κ of the second copy we have $V(\langle \kappa, 1\rangle) \models \neg\Diamond\ulcorner 0 \neq 0\urcorner$ because $\langle \kappa, 1\rangle$ cannot see the world $\langle 0, 0\rangle$, which is a dead end. Clearly both $\langle \kappa, 0\rangle$ and $\langle \kappa, 1\rangle$ can see themselves and are evaluated differently. ⊣

§7. Uniqueness of Valuations. Are there two distinct PW-models based on the same frame? That is, are there frames that admit distinct valuations? Converse wellfounded frames have exactly one valuation by Theorem 31; thus the valuation is always unique. Therefore two valuations for the same frame can differ only on converse illfounded worlds.

The technique developed in the previous section yields an example of a frame that admits two distinct valuations. In order to construct the two valuations, we need an illfounded model of KP.

LEMMA 35. *There is a nonwellfounded model of* KP *with*

$$On(\mathrm{WFP}(\mathcal{A})) = \kappa.$$

PROOF. This is a reasonably standard argument, but we explicitly give it here. For the unexplained notions the reader may consult [3]. By Theorem 23, L_κ is a union of KP-models: there are unboundedly many $\alpha < \kappa$ with $L_\alpha \models \mathrm{KP}$ (for otherwise if α is the largest such, κ is Σ_1-definable in $L_{\kappa+1}$ using the parameter α as "the least $\beta > \alpha$ so that $L_\beta \models$ KP"). Now let T be the following theory in the infinitary language $\mathcal{L}_{\kappa,\omega}$ with constant symbols $\bar{x}$ for each $x \in L_\kappa$ and a further named constant symbol $\bar{c}$. The theory T comprises the following sets of sentences:

(1) KP + $V = L$.
(2) The diagram of L_κ: $\{\, \bar{x} = \bar{y} \mid x = y \in L_\kappa \,\}$; $\{\, \bar{x} \in \bar{y} \mid x \in y \in L_\kappa \,\}$.
(3) For each $\beta < \kappa$ the sentence "$\bar{\beta} \leq \bar{c} \ \wedge\ \bar{c}$ is an ordinal".
(4) The sentence "$\forall v_0 \leq \bar{c}\ v_0$ is not the closure ordinal of Φ".

Then every $T_0 \subseteq T$ with $T_0 \in L_\kappa$ has a model: just take the least α with $L_\alpha \models \mathrm{KP}$ and $T_0 \in L_\alpha$, and interpret $\bar{c}$ as any ordinal $< \alpha$ larger than any ordinal mentioned in T_0, and interpret the other constant symbols occurring in T_0 as the corresponding actual sets. By the Barwise Compactness Theorem [3, Theorem 5.6], T has a model $\mathcal{A}$ which properly end extends L_κ by (2) and (3) (as $\bar{c}_{\mathcal{A}}$ is interpreted as something greater than all $\alpha < \kappa$); but κ is not in the wellfounded part of $\mathcal{A}$: otherwise it would be $\leq \bar{c}_{\mathcal{A}}$, and so there would be a closure point of Φ which is an initial segment of $On(\mathcal{A})$ below $\bar{c}_{\mathcal{A}}$, contradicting (4). ⊣

THEOREM 36. *There is a linearly ordered frame* $\langle W, R\rangle$ *which admits two different valuations* V, V'.

PROOF. To see this, let $\mathcal{A}$ be an admissible set whose wellfounded part has ordinal rank exactly κ. Such a set exists according to the previous lemma. The frame for which we construct two valuations V and V' is $\langle On(\mathcal{A}), <_{\mathcal{A}}\rangle$ where $On(\mathcal{A})$ is the set of ordinals of $\mathcal{A}$ (including the nonstandard ordinals) ordered in the sense of $\mathcal{A}$.

Let R^{-1} be an initial segment of the ordinals of $\mathcal{A}$ of length $b \in \mathrm{Ord}_{\mathcal{A}}$ where the latter is in the illfounded part of $\mathcal{A}$ and $\mathcal{A} \models b < \kappa_{\mathcal{A}}$ (if $\mathcal{A}$ thinks the latter exists). Define inside $\mathcal{A}$ the valuation V by recursion along $<_{\mathcal{A}}$. As $\kappa \notin \mathcal{A}$ the latter thinks that the closure ordinal $\kappa_{\mathcal{A}}$ (if it exists) is in the illfounded part of $\mathcal{A}$. Hence between κ and b there is no fixed point of the operator Φ. But now let V' be the valuation obtained by letting $V(w) = V'(w) = X_\alpha$ for w of rank $\alpha < \kappa$, and $V'(w) = X_\kappa$ otherwise. V

and V' are both valuations admitted by $\langle On(\mathcal{A}), <_{\mathcal{A}}\rangle$ (in the real world), but they are clearly different. $\dashv$

We conclude this section with a remark that actually belongs in section 5, for it complements Theorem 29. However, because we employ methods similar to those of the proof of Lemma 35, we have placed it here.

PROPOSITION 37. *For any countable* $\alpha \in$ ADM *there is a linearly ordered frame* $\langle W, R\rangle$ *with* R^{-1} *nonwellfounded, with* $\mathrm{WFP}(R^{-1}) \cong \alpha$ *and* $V(w)$ *is a fixed point for some* $w \in W$.

PROOF. It clearly suffices to assume that $\alpha < \kappa$. Apart from the last requirement of a fixed point, this would follow immediately from the standard construction of a nonwellfounded model $\mathcal{A}$ of KP with $\mathrm{WFP}(\mathcal{A}) \cong \alpha$—which can be effected by an omitting-types argument. We give a short proof here using a result of Friedman–Jensen, and of Sacks. This runs as follows:

THEOREM 38 (Friedman–Jensen; Sacks [40]). *Let* $\alpha \in$ ADM. *Then there is* $x \subseteq \omega$ *so that* $\omega_1^x = \alpha$.

Now let $\mathcal{A}_0 = L_\delta[x]$ the δ'th level of the constructible hierarchy using x as predicate. Let $\delta > \kappa$ be chosen least so that $L_\delta[x] \models$ KP. Consider the following theory T in the language $\mathcal{L}_{\alpha,\omega}$:
(1) KP + $V = L[\dot{x}]$ + "$\dot{x} \subseteq \omega$".
(2) The diagram of $L_\alpha[x]$:

$$\{\, \bar{x} = \bar{y} \mid x = y \in L_\alpha[x] \,\} \cup \{\, \bar{x} \in \bar{y} \mid x \in y \in L_\alpha[x] \,\}.$$

(3) For each $\beta < \alpha$ the sentence "$\bar{\beta} \leq \bar{c} \wedge \bar{c}$ is an ordinal".
(4) The sentence "$\forall v_0 \leq \bar{c} \bigvee_{\psi \in \mathrm{KP}} L_{v_0}[x] \models \neg\psi$".
(5) "The closure ordinal of Φ exists."
Then every $T_0 \subseteq T$ with $T_0 \in L_\alpha[x]$ has a model: choose $\eta < \alpha$ so that any β mentioned in T_0 from the collection specified at (3) is less than η. Then an interpretation of $\bar{c}$ is afforded by η, and thus $\langle L_\delta[x], \in, x, \eta\rangle$ is a model of T_0. (4) holds since η is less than the first ordinal that is the height of an admissible set containing x. The theory T is Σ_1-definable over $L_\alpha[x]$ and so the Barwise Compactness Theorem again applies: so let $\mathcal{A}$ be a model of T. Then we claim that $\mathrm{WFP}(On(\mathcal{A})) = \alpha$. By (1) and (2) $\mathcal{A} \supseteq L_\alpha[x]$. Hence $\mathrm{WFP}(On(\mathcal{A})) \geq \alpha$. But (3) ensures $\alpha \leq \bar{c}_{\mathcal{A}}$, and (4) that α cannot be in the wellfounded part of $\mathcal{A}$. $\dashv$

In particular, it follows that the minimal depth of an illfounded world in any model is exactly ω_1.

§8. Completeness. Theorem 32 imposes very strong limitations on frames that support PW-models. The examples in Section 3 illustrate the restrictions. Does this show that possible-worlds semantics does not reach far, that is, that there are no possible-worlds models where models ought to exist? In this section we show that our semantics is not to be blamed for these limitations because there is a PW-model for any set of sentences consistent in ω-logic.

In operator modal logic completeness theorems play a central role. For instance, the sentences valid in all frames at all words under any valuation are exactly the theorems of the modal system K. Similarly, the system $\mathsf{K4}$ is associated with the class of transitive frames, the Gödel system G with the class of converse wellfounded frames (see Boolos [6]).

Since we are dealing with standard models only, the set of all sentences valid in all PW-models $\langle W, R, V\rangle$ contains all arithmetical truths and is closed under the ω-rule. Of course this set is not recursively enumerable.

Definition 39. A set *Th* of sentences of $\mathcal{L}_\Box$ is $\Box$-*closed* if and only if it satisfies the following conditions:

(i) *Th* contains PA and is closed under logic, necessitation and the ω-rule:

$$\frac{A(\bar{0}), A(\bar{1}), \ldots}{\forall x\, A(x)}$$

(ii) $\Box\ulcorner A \to B\urcorner \to (\Box\ulcorner A\urcorner \to \Box\ulcorner B\urcorner)$ is in *Th* for all A and B in $\mathcal{L}_\Box$.

(iii) All instances $\forall x\, \Box\ulcorner A(\dot{x})\urcorner \to \Box\ulcorner\forall x\, A(x)\urcorner$ of the Barcan formula (i.e., the formalised ω-rule) are in *Th* for all formulae $A(v)$ of $\mathcal{L}_\Box$.

Theorem 40 (Completeness). *If Th is $\Box$-closed and consistent, then there is a PW-model $\langle W, R, V\rangle$ such that for all sentences $A \in \mathcal{L}_\Box$ the following equivalence holds: $A \in$ Th iff A holds in $\langle W, R, V\rangle$ at any world.*

Our completeness proof resembles the usual proofs for operator modal logic (see, e.g., Chagrov and Zakharyaschev [10, chapter 5]) by the method of canonical models. However, since we are dealing with standard models only, we employ ω-logic rather than pure first-order logic.

The worlds of the model constructed in the proof are just the possible extensions of the box. Thus one might think of the worlds as the standard models for *Th*.

For a set *Th* of sentences of $\mathcal{L}_\Box$ we write $(\mathbb{N}, S) \models Th$ if for all sentences $A \in \mathcal{L}_\Box$, $(\mathbb{N}, S) \models A$ holds.

Proof. Assume that *Th* is consistent and define W as the following set:

$$W =_{df} \{\, S \subseteq \omega \mid (\mathbb{N}, S) \models Th \,\}.$$

W is not empty since *Th* is consistent in ω-logic (see Barwise [3, Corollary 3.6]). Thus every world of W is a possible extension S of $\Box$ and we interpret $\Box$ at a world by S. We define the valuation V by setting $V(w) = w$. Therefore $(\mathbb{N}, V(w)) \models Th$ holds for all $w \in W$.

The accessibility relation R is defined in the following way:

$$wRv \;:\Longleftrightarrow\; (\mathbb{N}, v) \models w.$$

Here w is conceived as a set of sentences of $\mathcal{L}_\Box$.

It remains to check whether $\langle W, R, V\rangle$ is a PW-model, that is, we have to show

$$w \;=\; V(w) \;=\; \bigl\{\, A \in \mathcal{L}_\Box \;\big|\; \forall u(wRu \;\Rightarrow\; (\mathbb{N}, u) \models A) \,\bigr\}.$$

The left-to-right inclusion follows directly from the definition of R.

In order to prove the other direction, assume $\forall u(wRu \Rightarrow (\mathbb{N}, u) \models A)$. By inserting the definition of R we obtain the following:

$$\forall u\bigl((\mathbb{N}, u) \models w \;\Rightarrow\; (\mathbb{N}, u) \models A\bigr). \tag{13}$$

That is, A holds in all ω-models of w.

Since $(\mathbb{N}, w) \models \mathit{Th}$ and Th is $\Box$-closed, w contains all logical truths and all true atomic and negated atomic sentences of $\mathcal{L}_{\mathrm{PA}}$ by (i) above; w is closed under logic and the ω-rule because of $(\mathbb{N}, w) \models \mathit{Th}$ and (iii). This allows us to apply the ω-Completeness Theorem in the form of Barwise's [3, Theorem 3.5] $\mathfrak{M}$-Completeness Theorem to (13), which implies $A \in w$. ⊣

Corollary 41. *A $\Box$-closed set Th of sentences is consistent iff it is satisfied by the dead end model, in which there is just one world and the accessibility relation is empty.*

Proof. From left to right: If Th is consistent, it has a PW-model by the Completeness Theorem 40. But by Theorem 32 every frame supporting a PW-model has dead ends. Thus every dead end satisfies Th. The other direction is obvious. ⊣

From the Completeness Theorem we can also derive several inconsistency results. For instances, we can prove that the theory FS of truth in Halbach [18], which is equivalent to a system introduced by Friedman and Sheard [14], is inconsistent in ω-logic. Here we present another example.

$\mathcal{D}$ is the smallest $\Box$-closed set containing all instances of $\Box\ulcorner A\urcorner \to \Diamond\ulcorner A\urcorner$ for all $A \in \mathcal{L}_\Box$. Obviously $\mathcal{D}$ is the predicate analogue (plus the ω-rule) of the system D of deontic logic. Note that, according to the definition of $\mathcal{D}$, the ω-rule and the rule of necessitation may be applied to consequences of $\Box\ulcorner A\urcorner \to \Diamond\ulcorner A\urcorner$ for all $A \in \mathcal{L}_\Box$. The following example might also the viewed as a variant of McGee's theorem.

Example 42. $\mathcal{D}$ is inconsistent.

Proof. For every tautology $\top$ we can derive $\Diamond\ulcorner\top\urcorner$ in $\mathcal{D}$. But $\Diamond\ulcorner\top\urcorner$ is not satisfied in a dead end and thus $\mathcal{D}$ is inconsistent by our previous corollary. ⊣

We define the predicate counterparts of the operator modal systems K and $\mathsf{K4}$.

Definition 43. $\mathcal{K}$ is the smallest $\Box$-closed set. $\mathcal{K}4$ is the smallest $\Box$-closed set that contains all instances of $\Box\ulcorner A\urcorner \rightarrow \Box\ulcorner\Box\ulcorner A\urcorner\urcorner$.

By applying the theorem to $Th = \mathcal{K}$ and $Th = \mathcal{K}4$ we get the following two corollaries:

Corollary 44. *The following are equivalent for all* $A \in \mathcal{L}_\Box$*:*

(i) A is valid at all worlds in all PW-models, i.e., $V(w) \models A$ for all PW-models $\langle W, R, V\rangle$ and $w \in W$.
(ii) A is in $\mathcal{K}$.

Proof. In order to prove that (i) implies (ii), we apply our Completeness Theorem. If $A \notin \mathcal{K}$ there is a world w in the PW-model $\langle W, R, V\rangle$ constructed in the proof of the completeness result such that $A \notin w$, i.e., $V(w) \not\models A$.

The Normality Lemma 2.3 ensures that (ii) implies (i). ⊣

We obtain also an analogous result for $\mathcal{K}4$:

Corollary 45. *The following are equivalent for all* $A \in \mathcal{L}_\Box$*:*

(i) A is valid at all worlds in all transitive PW-models, i.e., $V(w) \vDash A$ for all transitive PW-models $\langle W, R, V\rangle$ and $w \in W$.
(ii) A is in $\mathcal{K}4$.

Not all completeness theorems known from operator modal logic carry over to their predicate counterparts. For instance, the logic G obtained from $\mathsf{K4}$ by adding the operator version $\Box(\Box p \rightarrow p) \rightarrow \Box p$ is complete with respect to the class of all transitive converse wellfounded frames. The predicate system $\mathcal{K}4$ has Löb's Theorem already as a theorem. However, it does not prove all sentences valid in all transitive converse wellfounded frames, as we shall show in the next section.

In possible-worlds semantics for operators of sentences one can build PW-models on all frames. We have proved that there are very strong limitations on possible-worlds semantics for predicates (Theorem 32). In particular, the frames of the modal systems T, $\mathsf{S5}$ and D do not feature dead ends and therefore they do not admit valuations. Thus the predicate approach seems to rule out the most important kinds of modality.

Does this not show that the operator approach is superior? We do not think so. It is not our possible-worlds semantics that is to be blamed. Obviously, the predicate analogues of T and $\mathsf{S5}$ are inconsistent, and the analogue of D becomes inconsistent if some natural closure conditions are added (Example 42). Thus the fact that there are no predicate PW-models based on T, $\mathsf{S5}$ and D mirrors only these inconsistency results. In general, the Completeness Theorem shows that there is a PW-model in our sense for any set of sentences that is consistent in the logic described in Definition 39.

Even if we generalise our present account and allow for nonstandard models, we will not be able to provide valuations for reflexive frames (see section 11). In general, our results make the limitations imposed by Löb's Theorem and its relatives explicit.

§9. The Fixed Point $\Phi^\kappa(\mathcal{L}_\Box)$. By Corollary 45, $\mathcal{K}4$ is the set of sentences valid in all transitive PW-models; $\Phi^\kappa(\mathcal{L}_\Box)$ is the set of all sentences valid in a all transitive *converse wellfounded* PW-models. Therefore $\Phi^\kappa(\mathcal{L}_\Box)$ is a subset of $\mathcal{K}4$. We shall show that the converse does not hold: there are sentences that are valid in all transitive converse wellfounded PW-models but that fail at a world of some transitive converse illfounded PW-model. In particular, Corollary 45 does not hold if (i) is restricted to converse wellfounded frames.

$\mathcal{K}4$ is obviously Π^1_1. The following theorem shows that $\Phi^\kappa(\mathcal{L}_\Box)$ has a much higher complexity than $\mathcal{K}4$. Thus converse illfounded PW-models really do matter.

THEOREM 46. *The set $\Phi^\kappa(\mathcal{L}_\Box)$ is Δ^1_2 but neither Σ^1_1 nor Π^1_1.*

PROOF. Φ is a Δ^1_1-operator; hence its fixed point $\Phi^\kappa(\mathcal{L}_\Box)$ is Δ^1_2 (see, for example, Hinman [22, Theorem 3.10]).

As (3) before Lemma 24 shows, $\Phi^\kappa(\mathcal{L}_\Box)$ is essentially the Σ_1-theory of a reasonably high countable level of the Gödel hierarchy, well beyond the first admissible level. Its complexity is thus well beyond Π^1_1. This will have been amply demonstrated to those familiar with the Gödel hierarchy, or the ramified analytical hierarchy. However, we can now give a direct demonstration of this fact.

In order to show that $\Phi^\kappa(\mathcal{L}_\Box)$ is not Π^1_1 we use the match between $X_\alpha = \Phi^\alpha(\mathcal{L}_\Box)$ and Σ_1-Th(L_α) established in Lemma 24.

We show that $\Phi^\kappa(\mathcal{L}_\Box)$ is at least Π^1_1 in a complete Π^1_1-set of integers (so of the degree $\mathcal{O}^\mathcal{O}$ where $\mathcal{O}$ is Kleene's ordinal notation system for ordinals up to ω_1. $\mathcal{O}^\mathcal{O}$ is Kleene's set of ordinal notations relativised to the set $\mathcal{O}$, and hence is a complete $\Pi^1_1(\mathcal{O})$-set of integers.

Let B be a sentence with $B \in X_{\omega_2+1} \setminus X_{\omega_2}$ where ω_2 is the second admissible ordinal.

We define the following recursive function f for Σ_1-sentences τ of the language of set theory:

$$f(\tau) := \big(C_1(\ulcorner\tau\urcorner) \wedge (C_1(\ulcorner\tau\urcorner) \leftrightarrow (\neg B \wedge \neg\Diamond\ulcorner B\urcorner))\big).$$

Put $f(n) = 0$ for other n that are not Σ_1-sentences of the language of set theory.

Then $f^{-1}{}''(X_\kappa)$ is the set Σ_1-Th(L_{ω_2}). The latter is known to be (recursively isomorphic to) $\mathcal{O}^{\mathcal{O}}$ (see [23]). Thus, $\mathcal{O}^{\mathcal{O}}$ is recursively reducible to X_κ. Hence X_κ cannot be Π^1_1, or else so would $\mathcal{O}^{\mathcal{O}}$. ⊣

The method for establishing the lower bound is general and can be applied also to admissibles greater than ω_2.

The previous theorem shows explicitly that $\Phi^\kappa(\mathcal{L}_\Box)$ is complex and hence cannot be equal to $\mathcal{K}4$. We can exhibit an explicit sentence that is true in all converse wellfounded worlds, but fails to be in $\mathcal{K}4$.

Let A be in $X_{\omega_1+1}\backslash X_{\omega_1}$ (any such will do for the proof). For the sake of definiteness we can obtain one as follows:

There is a Π_3-sentence τ in $\mathcal{L}_{\{\dot\in\}}$ so that $L_{\omega_1} \models \tau$ but for no smaller β does $L_\beta \models \tau$. For example it can be shown there are monotone total functions on ω cofinally increasing into ω_1 which are Σ_2-definable over L_{ω_1}; the statement that such a function is cofinal and defined on all of ω yields such a τ.

Thus τ is in the Π_3-Th(L_{ω_1}). Let $A = C_4(\ulcorner\tau\urcorner)$. Now let $\mathcal{A}$ be an illfounded KP-model of τ with WFP$(On(\mathcal{A})) = \omega_1$ (an adaptation of the argument of Proposition 37 gives us such an $\mathcal{A}$). Let $e \in \omega$ be an index for a recursive linear order (in fact a pseudo-wellorder) that $\mathcal{A}$ thinks is wellfounded, but is in fact illfounded. (Such an e must exist as τ is true in $\mathcal{A}$ only at $\mathcal{A}$'s first "admissible" ordinal, b say, so first at $(L_b)_\mathcal{A}$; $\mathcal{A}$ however thinks that b is the supremum of all recursive wellorders: hence there is an index e for a recursive linear ordering e which is isomorphic to an initial segment of $On(\mathcal{A})$ but is in fact illfounded.)

Let n be the code of the sentence expressing that $\{e\}$ is illfounded.

Theorem 47. $\Diamond\ulcorner A\urcorner \longrightarrow C_1(n)$ *is not provable in* $\mathcal{K}4$*, but it is true in all converse wellfounded worlds.*

Proof. In $\mathcal{A}$ we can define a model $\langle W, R, V\rangle$ and with $\mathcal{A}$ thinking that the frame is converse wellfounded and that $\Diamond\ulcorner A\urcorner$ is true in some world $w \in W$. (It suffices to have $\langle W, R^{-1}\rangle$ isomorphic to an initial segment of $On(\mathcal{A})$ containing b and some w with wRb.) However we could not have that $C_1(n)$ were true at w, as this would mean that $\mathcal{A}$ recognised e's illfoundedness—which it cannot. Thus $\mathcal{K}4$ does not prove the displayed sentence.

Let $\langle W, R, V\rangle$ be any model and w any converse wellfounded world in W with $\Diamond\ulcorner A\urcorner$ holding at w. L_{ω_1+1} realises that $\{e\}$ is truly illfounded: the wellfounded part of $\{e\}$ is mapped by its ranking function onto ω_1 by a Σ_1-recursion over L_{ω_1}. Hence the rest of the field of $\{e\}$—that is, its illfounded part, U say, is definable over L_{ω_1} and hence is a set in L_{ω_1+1}, and the latter is a model of "U has no $\{e\}$-minimal element" (a Σ_1-statement: $\forall k \in U\ \exists m \in U\big[\{e\}(\langle m, k\rangle) = 1\big]$—the bracketed matrix expressing that "m is $\{e\}$-below k"). Therefore $\Diamond\ulcorner A \urcorner \longrightarrow C_1(n)$ holds at w. ⊣

§10. Theories for $\Phi^\kappa(\mathcal{L}_\Box)$. Investigations into axiomatic theories based on PA plus an additional truth or necessity predicate axiomatised by suitable axioms turned out to be interesting and fertile (see, e.g., Cantini [9] for an overview). In many cases these theories are supposed to capture a semantical construction like Kripke's [26] theory. In this tentative section we explore the prospects of a strong theory for $\Phi^\kappa(\mathcal{L}_\Box)$.

We introduce some new notation. The quantifiers $\forall a$ and $\forall b$ range over sentences of the language $\mathcal{L}_\Box$; they can be defined in the usual way in $\mathcal{L}_{\mathrm{PA}}$. The quantifier $\forall a(v)$ ranges over all formulae of $\mathcal{L}_\Box$ with at most one free variable. The function sending the code of a formula to the code of its negation is expressed in $\mathcal{L}_{\mathrm{PA}}$ by $\underset{\cdot}{\neg}$, and similarly for other connectives and quantifiers. For the notation see also Feferman [11].

The following sentences are valid in $\Phi^\kappa(\mathcal{L}_\Box)$, that is, for any sentence A below we have $(\mathbb{N}, \Phi^\kappa(\mathcal{L}_\Box)) \models A$:

A1 All axioms of PA including all induction axioms in $\mathcal{L}_\Box$
A2 $\forall a\,(\mathrm{Bew}_{\mathrm{PA}}(a) \to \Box a)$
A3 $\forall a, b\,\big(\Box(a \underset{\cdot}{\to} b) \to (\Box a \to \Box b)\big)$
A4 $\forall a\,(\Box \underset{\cdot}{\Box} \dot{a} \leftrightarrow \Box a)$
A5 $\forall a\,(\Box \underset{\cdot}{\neg} a \to \neg \Box a)$
A6 $\forall a(v)\,\big(\Box \underset{\cdot}{\forall} v\, a(v) \leftrightarrow \forall x\, \Box a(\dot{x}/v)\big)$
A7 $\forall x\,\big(\Box \ulcorner A(\dot{x}) \urcorner \to A(x)\big)$ for all $A(v) \in \mathcal{L}_\Box$

The theory given by A1–A7 is at least as strong as ID_1. This follows from results proved by Friedman and Sheard [14].

By Theorem 46 the fixed point $\Phi^\kappa(\mathcal{L}_\Box)$ is more complex than the minimal fixed point of Kripke's construction (see Burgess [7, 8] for more on the complexity of truth-theoretic constructions). We are unable to say whether the theory A1–A7 as it stands is already a correspondingly stronger one than ID_1. The scheme A7 already represents the full reflection property of γ.

To find a theory that more aptly describes $\Phi^\kappa(\mathcal{L}_\Box)$ and has greater proof-theoretic strength than the simple axioms A1–A7 perhaps one has to add axioms expressing this reflection more fully. We can express that γ is a limit of Π_n-reflecting ordinals by adding the following. We let Form_n be a recursive one-place predicate enumerating the Π_n-formulae of $\mathcal{L}_\Box$ with one free variable. Define

$$\Phi_n\colon\ \forall x\, \forall a(v)\, \big(\mathrm{Form}_n(a) \to (\Box a(\dot{x}/v) \to a(x/v))\big).$$

Then define the following scheme:

$$(\mathrm{A8}_n) \qquad\qquad \forall b(\Box(\ulcorner \Phi_n \urcorner \underset{\cdot}{\to} \underset{\cdot}{\Box} b) \to \Box b).$$

This scheme is true in $\Phi^{\kappa}(\mathcal{L}_{\Box})$ also, for each n. Similarly one could express that γ is a limit of points, each of which was, for all n, a limit of Π_n-reflecting ordinals, but it is not clear whether this process carries us any further forward. It is also worth noting that at each world of a valuation along a converse wellorder shorter than κ, one may use Gödel numbers of sentences that become true at worlds accessible to it, as "labels" of those worlds, and indeed to reconstruct the whole of the ordering above any world, together with the valuations associated to those worlds.

§11. Nonstandard PW-Models. So far we have concentrated on models of the form $(\mathbb{N}, S)$ where $\mathbb{N}$ is the standard model of $\mathcal{L}_{\mathrm{PA}}$ and S is a set of standard numbers. As a consequence, the Barcan formula is valid in all worlds on our approach.

In the following two sections we sketch alternative approaches where some other model than $\mathbb{N}$ is used. While the next section is concerned with models for other languages than $\mathcal{L}_{\mathrm{PA}}$, we consider nonstandard models of $\mathcal{L}_{\mathrm{PA}}$ in the present section, more precisely, instead of $(\mathbb{N}, S)$ we employ certain models of the form $(\mathcal{M}, S)$ where $\mathcal{M}$ is a model of full arithmetic, i.e., of all arithmetical truths. Of course, there are further options, like arbitrary models of PA, but we will not explore them in the present paper.

By invoking nonstandard models we shall be able to construct valuations for the frame $\langle \omega, Suc \rangle$ of Example 9, which does not admit a valuation if the PW-model is based on the standard model $\mathbb{N}$. Thus we shall be able to construct valuations for frames without dead ends, whereas on the standard account all frames admitting valuations must have a dead end according to Theorem 32.

In a standard model $(\mathbb{N}, S)$ the extension S of $\Box$ is a set of numbers. If we pass to nonstandard models $(\mathcal{M}, S)$, we have to decide whether we shall admit nonstandard numbers in S as well. In a nonstandard model $\mathcal{M}$ the syntactical predicates expressible in $\mathcal{L}_{\mathrm{PA}}$ (like "x is a propositional tautology") apply to standard as well as to nonstandard elements of $\mathcal{M}$. Thus if we want to have sentences like "All propositional tautologies are necessary", i.e., $\forall a(\mathrm{Taut}(x) \rightarrow \Box x)$ as valid, then S must comprise also nonstandard elements of $\mathcal{M}$.

This poses a problem for the applicability of our Definition 2, which says that the extension of $\Box$ at a world w is just the set of sentences true at all worlds seen by w. If this condition is reformulated for the nonstandard case, we have to say what it means for a nonstandard sentence, i.e., a sentence in the sense of $\mathcal{M}$, to be true at a world. There are several ways of defining truth for nonstandard sentences. Here we use (a slight modification of) the notion of a satisfaction class (see Kaye [24] and Kotlarski [25] for overviews).

Let a model $\mathcal{M}$ be given. And let S_1 be a set of elements of $\mathcal{M}$. Then S_2 is a satisfaction class for $(\mathcal{M}, S_1)$ if and only if the following conditions are met:

(i) All elements of $\mathcal{M}$ which are true atomic sentences in the sense of $\mathcal{M}$ are in S_2. (These are the elements a such that $\mathcal{M} \models \mathrm{Tr}_0(\bar{a})$, where $\mathrm{Tr}_0(x)$ is a truth definition for atomic sentences of $\mathcal{L}_\Box$.)

(ii) S_2 commutes with all quantifiers and connectives in the sense of $\mathcal{M}$. For instance, the negation of a sentence in the sense of $\mathcal{M}$ is in S_2 if and only if the sentence itself is not in S_2.

According to results by Kotlarski, Krajewski and Lachlan [25, 28] and assuming that $\mathcal{M}$ is a countable nonstandard model of PA, there is a satisfaction class for $(\mathcal{M}, S_1)$ if and only if $(\mathcal{M}, S_1)$ is recursively saturated. Every countable recursively saturated model has uncountable many satisfaction classes.

The following is a modification of Theorem 2 of Leitgeb [30]:

THEOREM 48. *There is a model $\mathcal{M}$ elementarily equivalent to $\mathbb{N}$ and a function V on ω such that $V(n)$ is a satisfaction class for $(\mathcal{M}, V(n+1))$. Moreover all induction axioms hold in all models $(\mathcal{M}, V(n))$.*

V may be seen as a valuation for the frame $\langle \omega, Suc \rangle$ where a world n sees a world k if and only if $k = n + 1$. For every n, $V(n)$ contains all (standard) sentences true at all immediately preceding worlds, i.e., true in $(\mathcal{M}, V(n+1))$. In addition $V(n)$ is a satisfaction class for $(\mathcal{M}, V(n+1))$. In this sense $V(n)$ is a set comprising all $\mathcal{L}_\Box$-sentences in the sense of $\mathcal{M}$ that are true at all preceding worlds. However, $V(n)$ is only one of many such classes because there are many different satisfaction classes.

§12. Extensions to Other Languages and Ground Models. In this section we generalise our approach and the methods developed so far to languages extending $\mathcal{L}_\Box$.

Our results carry over to acceptable models in a straightforward way. Thus we can use other well-behaved models $\mathcal{M}$ (with the appropriate language $\mathcal{L}_\mathcal{M}$) instead of the standard model of arithmetic.

DEFINITION 49. Let $\mathcal{M} = \langle M, R, \ldots \rangle$ be any structure. Let $\kappa_\mathcal{M}$ be the least ordinal so that $L_{\kappa_\mathcal{M}}(\langle M, R, \ldots \rangle)$ is the first level of the relativised Gödel L-hierarchy built over M, using elements of M as *urelemente*, which is first-order reflecting.

The methods of Section 4 can be used to show the following. Let $\mathcal{M}$ be a countable acceptable structure (in the sense of [35], that is, essentially $\mathcal{M}$ has a definable coding scheme). Consider the class of frames where now $\mathbb{N}$ has been replaced by $\mathcal{M}$ at each world of the frame. One may show:

(i) The fixed point $\Phi^{\kappa_{\mathcal{M}}}(\mathcal{L}_\Box)$ occurs in sufficiently long converse well-founded frames $\langle W, R\rangle$ at the ordinal $\kappa_{\mathcal{M}}$.
(ii) We may look at admissible sets $\mathcal{A}$ over the structure $\mathcal{M}$ (much as Barwise does in [3], and take orderings $R \in \mathcal{A}$ with $\mathcal{A} \models$ "R^{-1} is wellfounded" and construct valuations V for the worlds of W.
(iii) The Characterisation Theorem 32 of the wellfounded parts of frames admitting valuations holds with κ replaced by $\kappa_{\mathcal{M}}$ and the class ADM replaced by ADM$(\mathcal{M})$—the class of ordinals admissible with respect to the structure $\mathcal{M}$.

In general the results of sections 4 & 5 go through *mutatis mutandis*.

The formulae of the language $\mathcal{L}_{\mathcal{M}}$ are interpreted in every world in the same way. In particular, in the special case $\mathcal{L}_{\mathcal{M}} = \mathcal{L}_{\mathrm{PA}}$ we have considered in the previous sections, the arithmetical vocabulary is interpreted at every world by the standard model. Also the domain stays constant throughout all worlds.

As announced above, we show how to extend our account to languages with contingent vocabulary whose interpretation varies from world to world. The domains of the worlds are allowed to differ as long as they contain the natural numbers. We shall still keep the interpretation of the arithmetical vocabulary fixed; it has thus the status of 'logical' vocabulary. The problems (e.g., the Church–Langford argument) connected with this assumption have been touched upon in the first section and will be discussed in another paper.

In order to expand our account to languages with contingent vocabulary, we redefine the language $\mathcal{L}_\Box$. In this section $\mathcal{L}_\Box$ is assumed to contain all arithmetical vocabulary, a relativizing predicate N, $\Box$ and arbitrary further constants, function and predicate symbols. $\mathcal{L}_{\mathrm{PA}}$ is the sublanguage of $\mathcal{L}_\Box$ containing only arithmetical vocabulary where all quantifiers are relativised to $\mathsf{N}x$. The relativization of the quantifiers is needed because the domains of our models may contain objects beyond the numbers.

A *standard model* for $\mathcal{L}_\Box$ (redefined in this way) is a model of $\mathcal{L}_\Box$ that interprets the sublanguage $\mathcal{L}_{\mathrm{PA}}$ of $\mathcal{L}_\Box$ in the standard way. Thus the standard model of arithmetic is a submodel of every standard model of $\mathcal{L}_\Box$ and $\mathsf{N}x$ applies in every standard model of $\mathcal{L}_\Box$ exactly to all (standard) numbers. A standard model of $\mathcal{L}_\Box$ also needs to fix the interpretation of the arithmetical functions applied to objects that are not numbers.

Definition 2 is then modified in the following way:

Definition 50. A *PW-model* is a triple $\langle W, R, V\rangle$ such that $\langle W, R\rangle$ is a frame and V assigns to every $w \in W$ a standard model of $\mathcal{L}_\Box$ such that the following condition holds:

$$V(w) \;=\; \left\{ A \in \mathcal{L}_\Box \;\middle|\; \forall u(wRu \;\Rightarrow\; V(u) \models A) \right\}.$$

Thus V does not only interpret $\Box$ at every world but it covers also the contingent vocabulary, i.e., the vocabulary not in $\mathcal{L}_{\mathrm{PA}}$.

If the notion of a PW-model is extended in this way, the main arguments of the paper still go through. In particular, our partial solutions of the Characterisation Problem and the Completeness Theorem 40 still apply. Valuations, however, are no longer unique because even converse wellfounded frames admit valuations that interpret the contingent vocabulary in different ways (and consequently also $\Box$).

The interpretation of the contingent vocabulary is not completely arbitrary. In a converse illfounded model there cannot be a sentence that is true at all converse wellfounded worlds and false at all converse illfounded worlds because Löb's Theorem rules out such valuations. Therefore there are restrictions on the interpretation of the contingent vocabulary.

In such PW-models the Barcan formula for arithmetical quantifiers, i.e., those restricted by $\mathsf{N}x$, will still obtain. In order to formulate the Barcan formula and its converse also for unrestricted quantifiers we would have to appeal to assignments and we would have to conceive $\Box$ as a binary predicate. In contrast to operator quantified modal logic, the converse Barcan formula does not drop out from our framework as a logical truth. Furthermore shrinking domains are allowed, i.e., a world may see another world with a smaller domain.

With contingent vocabulary we can express further conditions on frames. The following axiom forces infinite branching, that is, every world where it holds is either a dead end or sees infinitely many other worlds:

$$\Box\ulcorner\bot\urcorner \;\vee\; \forall x\big(\mathsf{N}x \rightarrow \Diamond\ulcorner G\dot{x} \wedge \forall y(Gy \;\rightarrow\; \dot{x}=y \vee \neg\mathsf{N}y)\urcorner\big).$$

G is here a predicate symbol distinct from $\Box$ and not in the arithmetical language.

We plan to treat canonicity and further classical topics of operator modal logic as applied to our framework in a future paper.

REFERENCES

[1] Peter Aczel and Wayne Richter, *Inductive definitions and reflecting properties of admissible ordinals*, ***Generalized recursion theory*** (Jens E. Fenstad and Peter Hinman, editors), North Holland, 1973, pp. 301–381.

[2] Nicholas Asher and Hans Kamp, *Self-reference, attitudes, and paradox*, ***Properties, types and meaning*** (Gennaro Chierchia, Barbara H. Partee, and Raymond Turner, editors), vol. 1, Kluwer, Dordrecht, 1989, pp. 85–158.

[3] Jon Barwise, ***Admissible sets and structures***, Perspectives in Mathematical Logic, Springer Verlag, Berlin, 1975.

[4] George Bealer, ***Quality and concept***, Clarendon Press, Oxford, 1982.

[5] Nuel Belnap and Anil Gupta, ***The revision theory of truth***, MIT Press, Cambridge, 1993.

[6] George Boolos, ***The logic of provability***, Cambridge University Press, Cambridge, 1993.

[7] John P. Burgess, *The truth is never simple*, ***The Journal of Symbolic Logic***, vol. 51 (1986), pp. 663–81.

[8] ———, *Addendum to 'The truth is never simple'*, ***The Journal of Symbolic Logic***, vol. 53 (1988), pp. 390–92.

[9] Andrea Cantini, ***Logical frameworks for truth and abstraction. an axiomatic study***, Studies in Logic and the Foundations of Mathematics, vol. 135, Elsevier, Amsterdam, 1996.

[10] Alexander Chagrov and Michael Zakharyaschev, ***Modal logic***, Oxford Logic Guides, Oxford University Press, Oxford, 1997.

[11] Solomon Feferman, *Reflecting on incompleteness*, ***The Journal of Symbolic Logic***, vol. 56 (1991), pp. 1–49.

[12] Solomon Feferman and Clifford Spector, *Incompleteness along paths in progressions of theories.*, ***The Journal of Symbolic Logic***, vol. 27 (1962), pp. 383–390.

[13] Hartry Field, *Disquotational truth and factually defective discourse*, ***The Philosophical Review***, vol. 103 (1994), pp. 405–452.

[14] Harvey Friedman and Michael Sheard, *An axiomatic approach to self-referential truth*, ***Annals of Pure and Applied Logic***, vol. 33 (1987), pp. 1–21.

[15] Giorgio Germano, *Metamathematische Begriffe in Standardtheorien*, ***Archiv für Mathematische Logik***, vol. 13 (1970), pp. 22–38.

[16] Dorothy Grover, Joseph Camp, and Nuel Belnap, *A prosentential theory of truth*, ***Philosophical Studies***, vol. 27 (1975), pp. 73–125.

[17] Anil Gupta, *Truth and paradox*, ***Journal of Philosophical Logic***, vol. 11 (1982), pp. 1–60.

[18] Volker Halbach, *A system of complete and consistent truth*, ***Notre Dame Journal of Formal Logic***, vol. 35 (1994), pp. 311–327.

[19] Volker Halbach, Hannes Leitgeb, and Philip Welch, *Possible worlds semantics for predicates*, ***Intensionality*** (Los Angeles) (Reinhard Kahle, editor), Association for Symbolic Logic, 2002, to appear.

[20] Joseph Harrison, *Recursive pseudo-well-orderings.*, ***Transactions of the American mathematical Society***, vol. 131 (1968), pp. 526–543.

[21] Hans G. Herzberger, *Notes on naive semantics*, ***Journal of Philosophical Logic***, vol. 11 (1982), pp. 61–102.

[22] Peter Hinman, ***Recursion theoretic hierarchies***, Springer, Berlin, 1978.

[23] Carl Jockusch and Stephen Simpson, *A degree theoretic definition of the ramified analytical hierarchy*, ***Annals of Mathematical Logic***, vol. 10 (1975), pp. 1–32.

[24] Richard Kaye, ***Models of Peano arithmetic***, Oxford Logic Guides, Oxford University Press, 1991.

[25] Henryk Kotlarski, Stanislav Krajewski, and Alistair Lachlan, *Construction of satisfaction classes for nonstandard models*, ***Canadian Mathematical Bulletin***, vol. 24 (1981), pp. 283–293.

[26] Saul Kripke, *Outline of a theory of truth*, ***Journal of Philosophy***, vol. 72 (1975), pp. 690–712.

[27] ———, *Is there a problem about substitutional quantification?*, ***Truth and meaning: Essays in semantics*** (Gareth Evans and John McDowell, editors), Clarendon Press, Oxford, 1976, pp. 325–419.

[28] Alistair Lachlan, *Full satisfaction classes and recursive saturation*, ***Canadian Mathematical Bulletin***, vol. 24 (1981), pp. 295–297.

[29] Hannes Leitgeb, *Theories of truth which have no standard models*, ***Studia Logica***, vol. 21 (2001), pp. 69–87.

[30] ———, *Truth as translation - part B*, ***Journal of Philosophical Logic***, vol. 30 (2001), pp. 309–328.

[31] Vann McGee, *How truthlike can a predicate be? A negative result*, ***Journal of Philosophical Logic***, vol. 14 (1985), pp. 399–410.

[32] ———, ***Truth, vagueness, and paradox: An essay on the logic of truth***, Hackett Publishing, Indianapolis and Cambridge, 1991.

[33] Richard Montague, *Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability*, ***Acta Philosophica Fennica***, vol. 16 (1963), pp. 153–67, Reprinted in [34, 286–302].

[34] ———, ***Formal philosophy: Selected papers of richard montague***, Yale University Press, New Haven and London, 1974, Edited and with an introduction by Richmond H. Thomason.

[35] Yiannis N. Moschovakis, ***Elementary induction on abstract structures***, Studies in Logic and the Foundations of Mathematics, no. 77, North-Holland and Elsevier, Amsterdam, London and New York, 1974.

[36] Arthur Pap, ***Analytische Erkenntnistheorie***, Springer, Wien, 1955.

[37] Michael Rathjen, *Proof theory of reflection*, ***Annals of Pure and Applied Logic***, vol. 68 (1994), pp. 181–224.

[38] Jim des Rivières and Hector J. Levesque, *The consistency of syntactical treatments of knowledge*, ***Theoretical aspects of reasoning about knowledge: Proceedings of the 1986 conference*** (Los Altos) (Joseph Y. Halpern, editor), Morgan Kaufmann, 1986, pp. 115–130.

[39] Hartley Rogers, ***Theory of recursive functions and effective computability***, McGraw–Hill Book Company, New York, 1967.

[40] Gerald E. Sacks, *Countable admissible ordinals and hyperdegrees*, ***Advances in Mathematics***, vol. 99 (1976), pp. 213–262.

[41] Paul Schweizer, *A syntactical approach to modality*, ***Journal of Philosophical Logic***, vol. 21 (1992), pp. 1–31.

[42] Brian Skyrms, *An immaculate conception of modality*, ***Journal of Philosophy***, vol. 75 (1978), pp. 368–387.

[43] Barry Hartley Slater, *Paraconsistent logics?*, ***Journal of Philosophical Logic***, vol. 24 (1995), pp. 451–454.

[44] Peter Strawson, *Truth*, ***Analysis***, vol. 9 (1949), pp. 83–97.

[45] Albert Visser, *Semantics and the liar paradox*, ***Handbook of philosophical logic*** (Dov Gabbay and Franz Günthner, editors), vol. 4, Reidel, Dordrecht, 1989,

pp. 617–706.

UNIVERSITÄT KONSTANZ
FACHGRUPPE PHILOSOPHIE
POSTFACH 5560 D 21
78434 KONSTANZ, GERMANY
and
DEPARTMENT OF LOGIC & PHILOSOPHY OF SCIENCE
UNIVERSITY OF CALIFORNIA
3151 SOCIAL SCIENCE PLAZA
IRVINE, CA 92697-5100, U.S.A.
Current address: Universität Göttingen, Philosophisches Seminar, Humboldtallee 19, 37073 Göttingen, Germany
E-mail: Volker.Halbach@uni-konstanz.de

UNIVERSITÄT SALZBURG
INSTITUT FÜR PHILOSOPHIE
FRANZISKANERGASSE 1
5020 SALZBURG, AUSTRIA
E-mail: Hannes.Leitgeb@sbg.ac.at

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRISTOL
BRISTOL BS8 1TW, ENGLAND
and
UNIVERSITÄT WIEN
INSTITUT FÜR FORMALE LOGIK
WÄHRINGER STRASSE 25
1090 WIEN, AUSTRIA
Current address: Universität Bonn, Mathematisches Institut, Beringstraße 6, 53115 Bonn, Germany
E-mail: p.welch@bristol.ac.uk