POSSIBLE WORLDS SEMANTICS FOR PREDICATES

VOLKER HALBACH, HANNES LEITGEB, AND PHILIP WELCH

Abstract. We develop possible worlds semantics for $\Box$ as a predicate rather than as an operator of sentences. The unary predicate symbol $\Box$ is added to the language of arithmetic (or an extension thereof); this yields the language $\mathcal{L}_{\Box}$. Every world in our possible worlds semantics is the standard model of arithmetic plus an interpretation of $\Box$. We investigate possible–worlds models where $\Box\ulcorner A\urcorner$ is true at a world w if and only if A is true in all worlds seen by w. The paradoxes exclude certain frames from being frames for $\Box$ as a predicate. We provide some sufficient and also some necessary conditions on frames that are allowed to act as frames for the predicate approach. Completeness results for certain infinitary systems corresponding to well known modal operator systems are established. We draw some conclusions concerning the current state of the predicate approach to modalities.

§1. Modalities as Predicates. Modalities like necessity and possibility, may be analysed logically in essentially two ways: either as predicates, or as operators. In the first case they are applied to singular terms, whereas in the second case they are applied to formulæ, but in both cases the application gives us new formulæ. Thus the distinction between the operator and the predicate conception of necessity is made on the syntactical level at first. Both conceptions are tied to certain semantics respectively. If "necessary" and "possible" are regarded as predicates, they are interpreted as properties of objects and a decision has to be made concerning what precisely they should be predicates of: syntactical entities like sentences, or contents of syntactical entities like propositions (let us ignore further options like utterances or mental objects). In either case, necessity and possibility are properties of such entities, or, perhaps, relations between such entities and further objects. If "necessary" and "possible" are regarded as operators, they do not express properties or relations like predicate and relation expressions; necessity does not apply to anything – much like the logical connectives or the quantifiers. In this

---

Received by the editors January 9, 2002.

The closure ordinal γ for the operator Φ was suggested to Volker Halbach by Peter Aczel and independently by Philip Welch, who later agreed to be a co-author of the paper and proved Theorem 4.8. The work of Volker Halbach was partially sponsored by the research group *Logic in Philosophy* at the Universities of Tübingen and Konstanz.

sense the operator conception of necessity is radically deflationary. Similar considerations apply not only to necessity but also to the notions of knowledge, belief, future and past truth, obligation and so on, which have been treated in analogous fashions as necessity.

There are various arguments in favour of, or against, each of these options of logical representation, but we shall not dwell on the pros and cons in more detail because this would require a careful analysis (see Bealer [4] and Halbach [16]). Instead we are going to concentrate on a particular problem affecting modalities if regarded as unary predicates of sentences, i.e., according to the extensionalist version of the predicate view from above. This position has been held and defended most notably by Quine [25, 24]. We think that there would be very good reasons for preferring this account if it was not threatened by inconsistency: Montague [22] was the first to notice that the predicate version of the acknowledged modal system T is inconsistent if it is joined with arithmetic. This is due to the existence of self-referential expressions, which is a consequence of Gödel's diagonalisation lemma, which is in turn derivable from even weak systems of arithmetic. From this Montague concluded:

> Thus if necessity is to be treated syntactically, that is, as a predicate of sentences, as Carnap and Quine have urged, then virtually all of modal logic [...] must be sacrificed.

Note that the intensionalist version of the predicate view on modality, i.e., where necessity is considered a predicate of intensional entities, is not necessarily affected by analogous problems. Propositions might e.g. be much more different from their syntactical counterparts than sometimes thought of, and therefore there might not be any self-referential propositions; perhaps self-referential sentences do not express propositions at all (for a recent account following that line of thought see Slater [31]). On the other hand, if propositions have a syntactical structure similar to sentences, and if every well-formed sentence expresses a proposition, then inconsistency strikes again (as has been pointed out by Asher and Kamp [2]).

According to the operator account of modality, there are simply no self-referential sentences rising from the use of modal expressions, since in order to be self-referential the sentences with operators would have to contain themselves as a proper part, which they certainly do not in the standard operator languages (below we shall make a remark on operator languages with additional devices like propositional quantification). Thus self-reference is abandoned from the start by suppressing referential clauses and by that move also the inconsistencies are avoided. This is one of the reasons why the majority of philosophers and logicians have opted to reject the predicate view altogether.

For the security from inconsistency, however, one has to pay a high price: quantification is not expressible in operator languages unless one adds further devices like propositional quantifiers. For instance, there is no way of expressing "All tautologies are necessary" in a single sentence of the operator language. If auxiliary devices like propositional quantification are employed, we expect that these additional devices reintroduce the difficulties of the predicate approach (Grim [12] argues for this view).

We doubt that the inconsistencies justify the general preference of the operator view over the predicate conception. We shall sketch some arguments for this claim but leave a thorough discussion of these matters to another paper.

First of all, by far not every system of modal logic has an inconsistent predicate counterpart. E.g. the predicate version of the minimal normal operator system K is indeed consistent. Secondly, des Rivières & Levesque [27] and Belnap & Gupta [5, p. 240ff] have shown that sentences of modal operator systems can be translated into sentences of corresponding and consistent predicate systems such that provability is preserved. Thus, roughly speaking, everything that can be said with operators can (consistently) be expressed with a predicate. Only if the provability of axiom *schemes* is to be preserved, troubles are emerging: we can have $\Box\ulcorner A\urcorner \rightarrow A$ 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 in the set of sentences which are modality-free, and therefore self-referential instantiations of the T-scheme are excluded.

Along the main stream of the operator conception of the modalities there has been some work on the predicate approach. One 'modality', however, has even been treated as a predicate by most authors, namely truth. The motivation for preferring the predicate approach in this case is obvious: the trivial modal system with the 'T-sentences' $\Box p \leftrightarrow p$ as axioms for $\Box$ will hardly qualify as a satisfying analysis of truth. Again we leave a discussion of the involved relations holding between the truth predicate, modal operators, and substitutional quantification to another paper. For our present purposes, however, it is important that several logicians have provided various models for axiom systems with predicates of sentences (or their codes). Friedman and Sheard [11], for instance, have studied a variety of systems built over an arithmetical base theory plus axioms for an additional unary (truth) predicate. Although the additional predicate is usually labelled as a truth predicate, some systems rather resemble systems for necessity or other typical modalities. We think that this applies, e.g., to one important derivative of the Friedman & Sheard systems, namely Cantini's [7] system VF. Therefore we surmise that some work on axiomatic theories of truth is also relevant for the predicate conception of other notions.

Truth theorists have investigated a variety of axiom systems (see, e.g., Cantini [8], Halbach [15] and Sheard [30] for overviews). Consistency proofs for these systems have been provided by model constructions. One aim of the present paper is a generalisation of these model constructions: we do no longer consider a single system and construct a model for this system. Rather we investigate a natural family of systems and provide a general pattern for generating models. And vice versa we aim at a generalisation of the inconsistency results. By showing that there is no model of a certain kind, we shall be able to conclude that a (semiformal – in the sense of Schütte [29],Sect.28) system is inconsistent or that a certain formal system does not have an ω-model.

Of course, we cannot hope to capture all the sophisticated model constructions found in the literature in a single approach, simply because they are too varied. Here we take some inspiration from the operator approach. Modal logicians do not investigate arbitrary modal systems (i.e. logic plus arbitrary axioms and rules for the modal operator $\Box$) but they restrict their focus on certain well behaved theories. The most common class of such well behaved theories is known as the set of normal modal systems (see, e.g., Boolos [6]). We follow exactly the same route for predicates. In this way we shall obtain results on the semantics of a comprehensive and important class of systems with a modality predicate.

Opting for possible world semantics in a more or less standard form implies certain restrictions. For instance, necessitation holds in all normal modal systems. Our possible world semantics for predicates will verify an analogous rule. Thus axiom systems which are not compatible with necessitation (like the Kripke-Feferman system Ref(PA) of Feferman [10]), which is also known as KF, fall outside the scope of our present paper. We want to claim by no means that these systems are less interesting or natural than 'normal' systems, but it makes sense to restrict attention to a class of well behaved system, that is, 'normal' systems first. Just as in the case of modal logic one finally can go on and explore non-normal systems and their semantics in a general framework.

Our approach does not only include a restriction to normal modal systems or, more exactly, to the predicate analogues of normal modal systems (see Lemma 2.3), but also a restriction to systems consistent in ω-logic. This restriction is brought about by our choice of the standard model of arithmetic as the basis for our model constructions. First of all this restriction serves the purpose of avoiding the intricacies of nonstandard models of arithmetic, but furthermore systems consistent in ω-logic are certainly quite natural. (However, there are natural and philosophically significant ω-inconsistent systems like the system FS mentioned below).

In our approach we want to stick as closely as possible to the common possible worlds semantics for operators, i.e. we use frames of worlds such that every world is a classical model. Since we shall apply the predicate $\Box$ to numerical codes of sentences, we assume that every world is actually a standard model of arithmetic which is only extended by assigning some extension to the Box – thus we restrict ourselves at first to just the arithmetical vocabulary plus a unary modal predicate. Later we are going to extend this language by non-arithmetical and non-logical vocabulary. Our treatment of necessity by means of a predicate extends to other modalities as well, but for the sake of simplicity we shall only speak of necessity and not enumerate all other modalities to which our approach may be applied.

On the semantical level, and as is common in modal logic, necessity is considered as truth in all accessible worlds, i.e., $\Box\ulcorner A\urcorner$ is true in w if and only if A is true in all worlds accessible from w, where accessibility is a binary relation (defined by the frame) on the class of worlds. At this point the fundamental difference between possible worlds semantics for modal operator and modal predicate languages shows up: according to the *operator* conception, the satisfaction clause for $\Box\ulcorner A\urcorner$ can be turned into a proper *definition*, but not so in the predicate case. If a frame is given and a function that assigns a model for the $\Box$-free language to any world, then truth or falsity for all sentences without $\Box$ extends to truth for all sentences with $\Box$ as an operator in the usual recursive way. Thus one can always build a model on any given frame if the operator conception is presupposed. But 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. There is also no way to provide a definition of the (possibly transfinite) complexity of a formula that would allow for a recursive definition of truth at a world. The satisfaction clause for modal sentences now works just as a condition on models, such that every model that meets the conditions counts as a proper instance of a possible worlds semantics for predicates. 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 due to self-referentiality. Montague's [22] Theorem, for instance, shows that one cannot build possible worlds models on reflexive frames (see Example 3.1 below) if necessity is expressed by 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 any world (and if some conditions are met which will be explained below) . Another example is provided by McGee's [21] theorem on ω-inconsistent systems. It is a direct consequence of this theorem that a frame consisting of the natural numbers ordered by the predecessor relation (so that

$0R1R2R3R\dots$) does not support a possible worlds model (see Example 3.2 below).

Our question from above may finally be stated in the following form: which frames support a possible worlds model, if $\Box$ is a predicate, and if the object language allows for self-referentiality? We call the problem of characterising the frames which support possible worlds models for predicates the *Characterisation Problem*.

A satisfying answer to the characterisation problem would exhibit a graph-theoretical description of the frames supporting possible worlds models. The restrictions imposed by theorems like those due to Montague and McGee should be consequences of such a characterisation result. A system with a modal predicate might thus be checked for consistency (or rather consistency in omega logic) by investigating what its corresponding possible worlds semantics would look like and whether the latter satisfies our characterisation criteria. In this way paradoxes and inconsistency results for modalities as predicates can be modelled in a systematic and informative way.

Our development of possible worlds semantics for predicates also yields some further insights: it sheds new light on hierarchies of metalanguages; some results of provability logic and modal logic show up in a different fashion; we can state the sound and complete systems for necessity as a predicate corresponding to certain types of frames; our treatment of transitive converse wellfounded frames can be seen as a variant of the revision theory of truth.

We regard this paper as a preliminary view on the issues sketched above. We also omit the more involved proofs, which we shall give in [17], but we outline some of the essential ideas.

§2. Formal Development. In this section we develop formally the account set out in the previous section.

$\mathcal{L}_{\mathrm{PA}}$ 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 formulæ and $\mathcal{L}_{\Box}$ with set of all formulæ that may contain $\Box$. We do not distinguish between expressions (like formulæ 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, but we shall only do so later, in order to keep the exposition of the main definitions and results as simple as possible. Once they have been established they can be carried over easily to languages with additional vocabulary.

We shall consider *standard* models only, i.e., worlds with the set of natural numbers in the domain and the standard interpretation of the arithmetical vocabulary. Hence models for $\mathcal{L}_\Box$ 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 such 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 as given, e.g., in Hughes and Cresswell [19, p. 37f.] Boolos [6].

DEFINITION 2.1. 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. w *sees* v if and only if and only if wRv; alternatively we say that w is below v.

DEFINITION 2.2. 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:

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

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.

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:

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

If we had to include further 'contingent' vocabulary, the valuation V had not only to interpret $\Box$ but also this additional vocabulary. Since we are only concerned with the language $\mathcal{L}_\Box$, V needs to interpret only $\Box$ at every world and therefore it needs to assigning only a set of sentences to every world, i.e., the extension of $\Box$ at the world.

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. For this problem is trivial: every frame admits a model 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 present 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 2.3 (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) $V(w) \models \Box\ulcorner A \rightarrow B\urcorner \rightarrow (\Box\ulcorner A\urcorner \rightarrow \Box\ulcorner B\urcorner)$

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

$$(1) \qquad \forall x \forall y \big(\mathrm{Sent}_{\mathcal{L}_\Box}(x) \wedge \mathrm{Sent}_{\mathcal{L}_\Box}(y) \rightarrow \Box(x \dot{\rightarrow} y) \rightarrow (\Box x \rightarrow \Box y)\big)$$

$\dot{\rightarrow}$ represents the function that yields, when applied to two codes of sentences, their material implication. $\mathrm{Sent}_{\mathcal{L}_\Box}(x)$ 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 like 1. This does not make a difference because of the use of standard models.

§3. Some Limitative Results. In order to illustrate the limitations imposed by the main condition in Definition 2.2, we present some examples of frames that do not support PW-models.

As has been noted in the introduction, several restrictions on the class of frames supporting a PW-model are known. These restrictions correspond to several well known inconsistency results.

Example 3.1 (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.

Montague's theorem is a generalisation of Tarski's theorem which says that the scheme $\Box\ulcorner A\urcorner \leftrightarrow A$ for all sentences $A \in \mathcal{L}_\Box$ is inconsistent with a theory proving diagonalisation. The possible worlds analogue of the scheme $\Box\ulcorner A\urcorner \leftrightarrow A$ is a frame with only one world that sees itself. Of course, such a frame does not admit a valuation, because the frame is reflexive, and thus Example 3.1 implies the possible worlds analogue of Tarski's theorem which states that a frame with one single reflexive world seeing itself does not support a PW-model.

The arguments showing that a frame does not admit a valuation proceed by diagonal arguments (in our last example by a Liar sentence).

McGee's main theorem in [21] on the ω-inconsistency of a certain theory of truth imposes a restriction on the existence of PW-models based on a non-transitive frame. Pred is the predecessor relation: for any natural number n it holds only between n and $n+1$.

$$\begin{tikzcd} \vdots \\ 2 \arrow[u] \\ 1 \arrow[u] \\ 0 \arrow[u] \end{tikzcd}$$

Example 3.2 (McGee, Visser). $\langle \omega, \mathrm{Pred}\rangle$ does not support a PW-model.

In order to prove this, one can adapt McGee's proof in [21]), which uses a certain self-referential sentence with quantification. Instead of McGee's theorem Visser's theorem [33] on illfounded hierarchies of languages can be used for deriving the above result. The relations between the different ω-inconsistency results and their implications is studied by Leitgeb [20].

Also Gödel's Second Incompleteness Theorem can be rephrased in terms of possible worlds semantics. One may prove an analogue of Löb's Theorem, which will be very useful in latter sections.

Transitivity yields the predicate analogue of the 4 axiom:

Lemma 3.3. *Let $\langle W, R, V\rangle$ be 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 3.4 (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:*

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

The proof is just the usual proof of Löb's Theorem with the provability predicate replaced by the primitive symbol $\Box$.

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 Löb's Theorem:

COROLLARY 3.5. *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 3.4 (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. $V(w) \models \Box\ulcorner\bot\urcorner$ holds if and only if w is a dead end. It follows:

EXAMPLE 3.6 (Gödel's Second Incompleteness Theorem). 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 4 axiom scheme holds. This suffices for a proof of the formalized incompleteness theorem and we obtain $V(w) \models \Box\ulcorner\bot\urcorner \vee \Diamond\ulcorner\Box\ulcorner\bot\urcorner\urcorner$. $\dashv$

In the sequel we shall provide a theorem (Theorem 6.1) that implies all the above results and does away with the need for all the different diagonal sentences.

§4. Inductive Definitions of Valuations. A world w is converse wellfounded in a frame $\langle W, R\rangle$ if and only if every subset of $\{v \mid wR^*v\}$ has an R-maximal element, where R^* is the transitive closure of R. A set of worlds is converse wellfounded if and only if all its elements are converse wellfounded. Consequently, a set of worlds is converse wellfounded if and only if it is wellfounded with respect to R^{-1}.

So far we did not show that there are any PW-models. As Belnap and Gupta [5, Theorem 6E.5] have shown, the extension of $\Box$ can easily be defined by recursion on R^{-1} in converse wellfounded frames. It applies also to frames that are not transitive.

THEOREM 4.1. *If $\langle W, R\rangle$ is converse wellfounded, then $\langle W, R\rangle$ admits exactly one valuation V.*

PROOF. The extension for $\Box$ is defined by transfinite recursion on R^{-1}.

Assume $V(v)$ has been defined for all v such that wRv. Then $V(w)$ is defined in the obvious way:

$$V(w) := \{A \in \mathcal{L}_\Box \mid \forall v(wRv \Rightarrow V(v) \models A\}$$

Clearly there is no choice in the definition of $V(w)$, which is thus uniquely determined. $\dashv$

As we shall show later, there are frames that admit more than one valuation. Of course, these frames are converse illfounded (i.e., not converse wellfounded) .

We turn to the transitive frames and investigate the following special case of Characterisation Problem: which transitive frames support PW-models? Throughout the rest of the section we assume that $\langle W, R\rangle$ is transitive.

In order to fit the present definition in the general framework of inductive definitions, it is convenient to define the extension of $\Diamond$ rather than the extension of $\Box$. The choice of $\Diamond$ as the basic predicate does not make a difference, because $\Box$ and $\Diamond$ are interdefinable in the following way:

$$\Box x :\Longleftrightarrow \neg\Diamond \dot{\neg} x$$
$$\Diamond x :\Longleftrightarrow \neg\Box \dot{\neg}(x)$$

We employ the following operation $^{\Diamond}$ turning an extension of $\Box$ into the corresponding extension of $\Diamond$ and vice versa:

$$X^{\Diamond} = \{A \in \mathcal{L}_{\Box} | \neg A \notin X\}$$

We have $X^{\Diamond\Diamond} = X$ for deductively closed X.

We define a satisfaction relation $\models_{\Diamond}$:

$$(\mathbb{N}, X) \models_{\Diamond} A :\Longleftrightarrow (\mathbb{N}, X^{\Diamond}) \models A$$

Thus $(\mathbb{N}, X^{\Diamond}) \models_{\Diamond} A$ holds if and only if $(\mathbb{N}, X) \models A$. Usually we shall write $Y \models_{\Diamond} A$ for $(\mathbb{N}, Y) \models_{\Diamond} A$.

Definition 4.2. We define an operator $\Phi : X \mapsto \Phi(X)$ for all sets of natural numbers.

$$\Phi(X) := X \cup \{A \in \mathcal{L}_{\Box} | (\mathbb{N}, X) \models_{\Diamond} A\}$$

Lemma 4.3. *Assume $\langle W, R, V\rangle$ is a PW-model based on a transitive converse wellfounded frame. If w has rank α, then $V(w)^{\Diamond} = \Phi^{\alpha}(\emptyset)$ holds for all ordinals α.*

The operator Φ is not monotone, that is, $X \subseteq Y$ does not generally imply $\Phi(X) \subseteq \Phi(Y)$. For instance $\neg\Diamond\ulcorner 0 = 0\urcorner$ is an element of $\Phi(\emptyset)$, but it fails to be an element of $\Phi(\{0 = 0\})$. 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 union $\bigcup_{\alpha<\lambda} \Phi^{\alpha}(X)$ of all previous stages.

The hierarchy of sets $\Phi^{\alpha}(\emptyset)$ of sentences for all ordinals α may be seen as a transitive version of revision semantics as introduced by Gupta [13] and Herzberger [18] (see Belnap and Gupta [5]). For finite levels we can model revision in our possible–worlds semantics in the following way. Consider the frame $\langle\omega, \mathrm{Suc}\rangle$ where Suc is the successor relation. This frame is converse wellfounded because there is no infinite chain

$$n + 1 \,\mathrm{Suc}\, n \,\mathrm{Suc} \ldots \mathrm{Suc}\, 1 \,\mathrm{Suc}\, 0$$

Now think of our $\Diamond$ as the truth predicate. Actually we could also take $\Box$. This does not make a difference because the extensions of $\Diamond$ and $\Box$ coincide in worlds that see exactly one other world. The only exception is 0. Taking $\Box$ as the truth predicate corresponds to starting with all sentences in the extension of the truth predicate at level 0, while taking $\Diamond$ corresponds to starting with the empty set. In the revision semantics the *revision operator* yields applied to a set X the set of sentences that are true if the truth predicate is interpreted by X. Thus in our notation the revision operator Γ may be rephrased as

$$A \in \Gamma(X) \;:\Longleftrightarrow\; X \models_\Diamond A$$

In contrast to the situation with Φ, $X \subseteq \Gamma(X)$ does not hold. The liar sentence, for instance, will be in $\Gamma^n(\emptyset)$ if n is odd and it will not be in any $\Gamma^{2n}(\emptyset)$. Thus the sequence $(\Gamma^n(\emptyset))_{n\in\omega}$ of sets of sentences does not converge, while $(\Phi^n(\emptyset))_{n\in\omega}$ converges to $\Phi^\omega(\emptyset) = \bigcup_{n<\omega}\Phi^n(X)$. We state this important property for further reference as a lemma.

LEMMA 4.4. *If $\alpha \leq \beta$, then $\Phi^\alpha(\emptyset) \subseteq \Phi^\beta(\emptyset)$. Thus there is an ordinal κ such that $\Phi^\kappa(\emptyset) = \Phi^\alpha(\emptyset)$ for all $\alpha \geq \kappa$.*

The definition of the limit stages $\Gamma^\lambda(\emptyset)$ for the revision operator is the central controversial problem of revision semantics. If we force $(\Gamma^n(\emptyset))_{n\in\omega}$ to converge by taking not $\Gamma(x)$ but $\Gamma(X) \cup X$ at every step we obtain our operator Φ.

This kind of situation is encountered also with other operators Ψ producing sequences of sets that do not converge. Of course, one can always force convergence by passing from X not to $\Psi(X)$ but to $X \cup \Psi(X)$. If this operation is iterated, convergence is guaranteed as long as the image of the function Ψ is a set.

A set $X \subseteq \mathcal{L}_\Box$ is a fixed point (of Φ) if and only if $\Phi(X) = X$.

Since Φ is not monotone, $\Phi^\kappa(\emptyset)$ is not necessarily the smallest fixed point.

REMARK 4.5. $X \subseteq \mathcal{L}_\Box$ is a fixed point if and only if $X \models_\Diamond \Box\ulcorner A\urcorner \to A$ (or, equivalently, $X \models_\Diamond A \to \Diamond\ulcorner A\urcorner$) for all $A \in \mathcal{L}_\Box$.

The existence of fixed points implies that there are frames that are converse illfounded but which admit a valuation nevertheless, as we shall show in the following theorem.

In order to state the theorem, we need some definitions. The *wellfounded part* of a set of the form $\{v|\, wRv\}$ is 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 the rank of R^{-1} restricted to the converse wellfounded part of $\{v|\, wRv\}$. Obviously the depth is defined also for converse illfounded worlds. We shall apply these notions only to transitive frames.

THEOREM 4.6. *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 4.3. If $w \in W$ has depth at least κ we put $V(w) = \Phi^{\kappa}(\emptyset)$. This holds in particular for all converse illfounded worlds.

By Lemma 4.4 the condition

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

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

This and Löb's Theorem imply the following corollary:

COROLLARY 4.7. *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 $\langle W, R\rangle$.*

Using an ordinal notation system for the ordinals below ω_1^{CK}, the first non-recursive ordinal (see Rogers [28]), it is possible to construct sentences that hold at a world w of a transitive frame if and only if the world is converse wellfounded and w has depth $\alpha < \omega_1^{\mathrm{CK}}$. However, in general one cannot express converse wellfoundedness in $\mathcal{L}_{\Box}$ according to Corollary 4.7.

In the following we determine the exact size of the closure ordinal κ of Φ and thereby show that there is a gap between ω_1^{CK} and κ.

We abbreviate by setting $X_{\alpha} = \Phi^{\alpha}(\emptyset)$. The function $\delta \mapsto X_{\delta}$ can be defined by a Σ_1-recursion over any sufficiently closed model $\mathcal{M}$ 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. Models of KP are called admissible sets. An ordinal is admissible if L_{α} is a model of KP. ADM is the set of all admissible ordinals greater than ω.

The idea of the proof of the following theorem is that below the closure ordinal κ of this operation, the sets X_{δ} 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. Then one may show that $L_{\gamma} \prec_{\Sigma_1} L_{\gamma+1}$ and that 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 fails first at $L_{\gamma+1}$. (For background on the theory KP and on admissible sets, see Barwise [3].)

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]):

$$L_\gamma \models \varphi \implies \exists \alpha < \gamma \, L_\alpha \models \varphi$$

The proof theoretic strength of this ordinal lies strictly between ACA + Bar Induction + Δ^1_2-Comprehension Scheme, and the latter strengthened with the Π^1_2-Comprehension Scheme — see Rathjen [26].

THEOREM 4.8. *$\gamma = \kappa$.*

In particular, this shows that $\kappa > \omega_1^{\mathrm{CK}}$. Thus the characterisation of the class of transitive frames admitting valuations yielded by Lemma 6.1 below and Theorem 4.6 is incomplete because it does not say whether worlds of depth α with $\omega_1^{\mathrm{CK}} \leq \alpha \leq \kappa$ have to be converse wellfounded. We shall now show that these worlds may be illfounded, too.

For $\mathcal{A}$ an admissible set we let $\mathcal{WFP}(\mathcal{A})$ denote the *wellfounded part* of $\mathcal{A}$, that is, the largest structure $\mathcal{B}$ such that $\mathcal{A}$ is an end extension of $\mathcal{B}$. If α is the ordinal rank of the wellfounded part $\mathcal{WFP}(\mathcal{A})$, then it is known that $L_\alpha \models \mathrm{KP}$. If $\mathcal{A}$ is a model of KP, $On(\mathcal{A})$ denotes the class of ordinals in the sense of $\mathcal{A}$. We identify the standard part of $On(\mathcal{A})$ with the standard ordinals. Hence $On(\mathcal{WFP}(\mathcal{A})) \supseteq \omega + 1$ holds if and only if the integers of $\mathcal{A}$ are standard. We shall only consider in what follows admissible sets in which the integers are standard.

We can prove the following existence theorem for valuations, which yields valuations for converse illfounded frames if it is applied to nonstandard models $\mathcal{A}$ of KP.

THEOREM 4.9. *If $\langle W, R\rangle$ is transitive, $\mathcal{A}$ is a model of* KP*, $R \in \mathcal{A}$ and "every converse illfounded world in $\langle W, R\rangle$ has at least depth κ" holds in $\mathcal{A}$, then $\langle W, R\rangle$ admits a valuation.*

The statement of this theorem is not completely precise because we have to say how the condition in quotation marks is expressed in the language of set theory and how the relation R is represented. This may be done in a standard way and we do not go into the details.

§5. A Necessary Condition for the Existence of a Valuation. As above, we let ADM be the class of admissible ordinals without ω, that is, the class of ordinals $\alpha \neq \omega$ so that $L_\alpha \models \mathrm{KP}$. ADM^* is the class of ADM together with its limit points.

As we have explained above, we cannot prove that a frame must be converse wellfounded if it admits a valuation. However, Löb's Theorem can be employed for showing the following:

THEOREM 5.1. *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 $\alpha \geq \kappa$.*

Theorem 5.1 and Theorem 4.6 yield the following incomplete characterisation of the class of transitive frames that admit valuations: above depth ω_1^{CK} (the first non-recursive ordinal) all worlds have to be wellfounded, while below κ everything, i.e., arbitrary loops and infinitely descending chains are possible. Thus for depths smaller than ω_1^{CK} and 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 model must be admissible or a limit of admissibles according to Theorem 5.1. In the next proposition we shall see that there actually are models with converse illfounded worlds of depth ω_1^{CK}. Thus converse illfounded worlds occur as low as possible by Theorem 5.1. However, we are not able to provide exhaustive information on the converse illfounded worlds between ω_1^{CK} and κ.

In the following proposition $\mathcal{WFP}(R^{-1})$ denotes the largest wellfounded initial segment of R^{-1}.

Proposition 5.2. *For any $\alpha \in \mathrm{ADM}$ there is a frame $\langle W, R\rangle$ with R^{-1} illfounded, with $\mathcal{WFP}(R^{-1}) \cong \alpha$ and $V(w)$ is a fixed point for some $w \in W$.*

This follows from Barwise's [3] Compactness Theorem and the fact that for any countable ordinal $\alpha \in \mathrm{ADM}$ there is a illfounded model with a wellfounded initial part of order type α. We omit the proof.

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

We conclude this section with a note on the uniqueness of valuations.

Are there two distinct PW-models based on the same frame? That is, are there frames that admit at least two distinct valuations? Converse wellfounded frames have exactly one valuation by Theorem 4.1; thus the valuation is always unique. Therefore two valuations for the same frame can differ only on converse illfounded worlds. One can show the following by employing illfounded models of KP.

Theorem 5.3. *There is a linearly ordered frame $\langle W, R\rangle$ which admits two different valuations V, V'.*

§6. The General Characterisation Problem. So far we have focused on transitive frames. In this section we generalise some results to other frames.

Theorem 5.1 yields information also on frames that are not transitive. For the transitive closure of R can be defined within the language $\mathcal{L}_\Box$.

Theorem 6.1. *If $\langle W, R\rangle$ supports a PW-model, then in the transitive closure of $\langle W, R\rangle$ every converse illfounded world in W has depth α with $\alpha \in \mathrm{ADM}^*$ or $\alpha \geq \kappa$.*

Theorem 6.1 can be used for demonstrating that the frame $\langle\omega, \mathrm{Pred}\rangle$ in Example 3.2 does not admit a valuation. Thus Theorem 6.1 implies all negative results in § 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 § 7 we shall show how to get from these analogues back to ω-inconsistency results.

By Theorem 4.6 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 4.6 by showing that a frame admits a valuation if all converse illfounded 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.

THEOREM 6.2. *There are frames that do not support a PW-model, although their transitive closure does.*

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 $\Phi^{\kappa}(\emptyset)$. However, other fixed points of Φ are required for valuations of certain frames that are not transitive. There is a model with two reflexive worlds v and w such that $V(v) \neq V(w)$. Again we omit the proofs.

§7. Completeness. In 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]). Can we obtain analogous results in our setting?

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 7.1. 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:
$$\frac{A(\overline{0}), A(\overline{1}), A(\overline{2}), \dots}{\forall x A(x)}$$
(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 formulæ $A(v)$ of $\mathcal{L}_\Box$.

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.

THEOREM 7.2 (Completeness). *If Th is $\Box$-closed and consistent, then there is a PW-model $\langle W, R, V\rangle$ such that $(\mathbb{N}, V(w)) \models Th$ holds for all $w \in W$.*

Our completeness proof resembles the usual proofs for operator modal logic (see, e.g., Chagrov and Zakharyaschev [9, chapter 5]). However, since we are dealing with standard models only, we have to employ ω-logic rather than pure first-order logic.

From the Completeness Theorem we can also derive several inconsistency results. For instances, we can prove that the truth theory FS in Halbach [14], which is equivalent to a system introduced by Friedman and Sheard [11], is inconsistent in ω-logic. Here we present still 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 $A \in \mathcal{L}_\Box$.

EXAMPLE 7.3. $\mathcal{D}$ is inconsistent.

This follows from the Completeness Theorem and the fact that every frame admitting a PW-model contains a dead end, which contradicts the axiom scheme $\Box\ulcorner A\urcorner \to \Diamond\ulcorner A\urcorner$ of $\mathcal{D}$.

If the ω-rule is dropped from $\mathcal{D}$ the system becomes consistent (though ω-inconsistent because of McGee's [21] result) because it is a subtheory of the consistent system FS of Halbach [14].

We define the predicate counterparts of the operator modal systems K and K_4.

DEFINITION 7.4. $\mathcal{K}$ is the smallest $\Box$-closed set. $\mathcal{K}_4$ is the smallest $\Box$-closed set that contains all instances of $\Box\ulcorner A\urcorner \to \Box\ulcorner\Box\ulcorner A\urcorner\urcorner$.

By applying the theorem to $Th = \mathcal{K}$ and $Th = \mathcal{K}_4$ we get the following two corollaries:

COROLLARY 7.5. *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 7.6. *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., $V(w) \models A$ for all 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 K_4 by adding the operator version $\Box(\Box p \to p) \to \Box p$ 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 now by recursion-theoretic considerations, more precisely by comparing the complexities of $\mathcal{K}_4$ and the fixed point $\Phi^\kappa(\emptyset)$.

By Corollary 7.6 $\mathcal{K}_4$ is the set of sentences valid in all transitive PW-models; $\Phi^\kappa(\emptyset)^\diamond$ is the set of all sentences valid in a all transitive *converse wellfounded* PW-models. Therefore $\Phi^\kappa(\emptyset)$ 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 7.6 does not hold if (i) is restricted to converse wellfounded frames.

$\mathcal{K}_4$ is obviously Π^1_1. The following Theorem shows that $\Phi^\kappa(\emptyset)$ has a much higher complexity than $\mathcal{K}_4$. Thus converse illfounded PW-models really do matter.

Theorem 7.7. *The set $\Phi^\kappa(\emptyset)$ is Δ^1_2 but neither Σ^1_1 nor Π^1_1.*

§8. 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. In this case $\Box$ has to be conceived as a relation obtaining between formulas and variable assignments. Thereby we introduce genuine de re–modality.

Definition 8.1. Let $\mathcal{M} = \langle M, R, \ldots\rangle$ be any structure. Let $\kappa_\mathcal{M}$ be the least ordinal so that $L_{\kappa_M}(\langle M, R, \ldots\rangle)$ 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 § 4 can be used to show the following. Let $\mathcal{M}$ be a countable acceptable structure (in the sense of [23], 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 $\Phi^{\kappa_\mathcal{M}}(\emptyset)$ occurs in sufficiently long converse wellfounded 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 6.1 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 § 4 and § 5 go through *mutatis mutandis.*

The formulæ 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 does not allow for any variation in its evaluation, at least not on our account: it is interpreted in every world by the standard model.

Moreover, we could allow the universe to vary from world to world, as long as the codes of sentences and codes of sequences of all objects existing in the respective world are in the universe. However, we do not go that far afield from our original approach and we illustrate the effects of contingent vocabulary by considering a slight modification of our approach with $\mathcal{L}_{\mathrm{PA}}$. For our simple examples we need only an extension of $\mathcal{L}_{\mathrm{PA}}$ by an additional propositional parameter.

Let $\mathcal{L}^p_{\mathrm{PA}}$ be the language $\mathcal{L}_{\mathrm{PA}}$ of arithmetic extended by an additional propositional variable p. By adding $\Box$ we obtain the language $\mathcal{L}^p_{\Box}$. A PW-model is defined as in Definition 2.2 except that V has also to tell whether p is true at that world. More precisely $V(w)$ is a pair $\langle S, k\rangle$ where $S \subseteq \mathcal{L}^p_{\Box}$ is the extension of $\Box$ and k is the truth value 0 or 1 of p at w. If $V(w) = \langle S, k\rangle$, S has to satisfy the following condition:

$$S = \{A \in \mathcal{L}^p_{\Box} | \forall u(wRu \Rightarrow V(u) \models A\}$$

Of course $V(u) \models A$ for $A \in \mathcal{L}^p_{\Box}$ is defined in the obvious way.

In the following theorem we show that we are not completely free to assign truth values to p if the frame is converse illfounded.

Proposition 8.2. *Assume $\langle W, R\rangle$ supports a PW-model. Then the transitive closure of a frame $\langle W, R\rangle$ is converse wellfounded if and only if it allows for arbitrary interpretations of p.*

By the latter we mean that given a frame and given a function $g : W \to \{0, 1\}$ we have a model $\langle W, R, V\rangle$ based on that frame such that for all $w \in W$ and some $S \subseteq \mathcal{L}^p_{\Box}$ the equation $V(w) = \langle S, g(w)\rangle$ holds.

Proof. Assume that the frame is transitive.

Löbs Theorem 3.4 applies also to p, that is, the following holds for any valuation V and world w:

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

By stipulating $V(w) \models p$ for converse wellfounded w and $V(w) \models \neg p$ for converse illfounded w, one can exclude converse illfounded worlds for transitive frames (see Boolos [6, p. 75f] and Corollary 4.7).

Since the transitive closure of R can be defined in $\mathcal{L}_\Box$, the argument can be generalised to non-transitive frames. $\dashv$

Although we are not completely free to evaluate p in any world of a converse illfounded model in any arbitrary way, we can use p in order to express certain features of the frame (for instance, we can force a frame to branch). Here we do not explore what can be expressed with additional vocabulary, but it should be obvious that interesting conditions are expressible with more propositional parameters and by quantifying over these parameters.

If we allow also for additional objects in every world beyond the natural numbers and we change our setup accordingly, we shall encounter the usual problems of quantified modal logic. In particular, the Barcan formula will fail if we allow for increasing domains etc.

Each of these extensions is worth to be studied and philosophically relevant – indeed these extensions are more relevant than the setting that we have followed: it is certainly not satisfying, from a philosophical point of view, to concentrate on worlds which are standard models of arithmetic with an additional interpretation of the necessity predicate. But independent of what language is chosen and what the set of worlds looks like semantically, if a modality is to be expressed by a predicate, then every world has to have an interpreted theory of syntax in the background, and that means that every world has to comprise arithmetic (modulo coding). Thus, we have actually concentrated on the minimal prerequisites of a possible worlds semantics for predicates, and that is why our results also apply to the philosophically more relevant cases.

§9. Some Preliminary Conclusions. In this final section we try to draw some conclusions from our results.

We have shown how all limitative results of § 3 may finally be reduced to a single limitative result. The possible worlds analogues of Tarski's, Montague's, McGee's, Visser's theorems, which have been mentioned in § 3, as well as the analogues of some results not mentioned here like those by Thomason [32] and partially those by Friedman and Sheard [11] may be reduced to the single condition on frames in Theorem 6.1. Löb's Theorem is the origin of all the inconsistency and ω-inconsistency results. On the semantical side this is mirrored by the condition that any converse illfounded world must have a large depth; this condition excludes all simple loops, reflexive frames etc. from being frames that admit a PW-model. Thus we conclude that the mentioned paradoxes flow from a common source.

Our semantical investigations underline the central role of converse wellfoundedness of frames. Provability logic (see, e.g., Boolos [6]) thus appears in a new light, at least to us. The main results of provability logic might be understood as suggesting that there is a profound relation between provability and converse wellfoundedness. In contrast, our results show that, if one sets up possible worlds semantics for predicates, converse wellfoundedness always plays an important role. All frames supporting a PW-model have to be converse wellfounded — except for worlds of great depth. Thus all predicates with reasonable possible worlds semantics are much like the provability predicate. This holds even in the absence of the transitivity axiom as Theorem 6.1 shows.

Surprisingly the condition of Theorem 6.1 does not rule out all converse illfounded models. As Theorem 7.7 shows, our completeness result Corollary 7.5 for $\mathcal{K}$ would fail, if we restricted our attention to converse wellfounded models only. That is, $\mathcal{K}$ is incomplete with respect to the class $\Phi^{\kappa}(\emptyset)^{\diamond}$ of all converse wellfounded models. Consequently there are sentences that hold in all converse wellfounded models but fail at some converse illfounded model and are therefore not theorems of $\mathcal{K}$. This shows that the common semantical source of the paradoxes is not precisely converse illfoundedness of the frames but rather the failure of the more complex condition stated in Theorem 6.1.

The condition of Theorem 6.1 implies that many frames familiar from operator modal logic do not support PW-models. In particular, the frames for T, S_4 and S_5, i.e., all reflexive frames do not admit a valuation. It seems that we have to sacrifice the core systems of modal logic for the predicate approach. We do not conclude from this, however, that the predicate approach is mistaken altogether. We recommend to approach necessity and other notions much like truth in order to ban inconsistency from these notions. For truth a wide array of semantical and axiomatic approaches has been offered in the literature. For necessity there should be even more.

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] ANDREA CANTINI, *A theory of truth formally equivalent to ID_1*, ***The Journal of Symbolic Logic***, vol. 55 (1990), pp. 244–259.

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

[9] ALEXANDER CHAGROV and MICHAEL ZAKHARYASCHEV, ***Modal logic***, Oxford Logic Guides, Oxford University Press, Oxford, 1997.

[10] SOLOMON FEFERMAN, *Reflecting on incompleteness*, ***The Journal of Symbolic Logic***, vol. 56 (1991), pp. 1–49.

[11] HARVEY FRIEDMAN and MICHAEL SHEARD, *An axiomatic approach to self-referential truth*, ***Annals of Pure and Applied Logic***, vol. 33 (1987), pp. 1–21.

[12] PATRICK GRIM, ***The incomplete universe: Totality, knowledge, and truth***, MIT Press, Cambridge (Mass.), 1991.

[13] ANIL GUPTA, *Truth and paradox*, ***Journal of Philosophical Logic***, vol. 11 (1982), pp. 1–60.

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

[15] ———, *Truth and reduction*, ***Erkenntnis***, vol. 53 (2000), pp. 97–126.

[16] ———, *Semantics and deflationism*, ***Habilitationsschrift***, Universität Konstanz, 2001.

[17] VOLKER HALBACH, HANNES LEITGEB, and PHILIP WELCH, *Possible worlds semantics for modal notions conceived as predicates*, in preparation.

[18] HANS HERZBERGER, *Notes on naive semantics*, ***Journal of Philosophical Logic***, vol. 11 (1982), pp. 61–102.

[19] G. E. HUGHES and M. J. CRESSWELL, ***A new introduction to modal logic***, Routledge, London and New York, 1996.

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

[21] VANN MCGEE, *How truthlike can a predicate be?*, ***Journal of Philosophical Logic***, vol. 14 (1985), pp. 399–410.

[22] RICHARD MONTAGUE, *Syntactic treatments of modality, with corollaries on reflexion principles and finite axiomatizability*, ***Acta Philosophica Fennica***, vol. 16 (1963), pp. 153–167.

[23] YIANNIS MOSCHOVAKIS, ***Elementary induction on abstract structures***, North-Holland, Amsterdam, 1974.

[24] WILLARD VAN ORMAN QUINE, *Three grades of modal involvemen*, ***The ways of paradox***, Harvard University Press, Cambridge, Mass., revised and enlarged ed., 1976, pp. 158–176.

[25] ———, *Intensions revisited*, ***Contemporary perspectives in the philosophy of language*** (French, Uehling, and Wettstein, editors), University of Minnesota Press, Minneapolis, 1977, pp. 5–11.

[26] MICHAEL RATHJEN, *Proof theory of reflection*, ***Annals of Pure and Applied Logic***, vol. 68 (1994), pp. 181–224.

[27] 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.

[28] HARTLEY ROGERS, ***Theory of recursive functions and effective computability***, McGraw–Hill Book Company, New York, 1967.

[29] Kurt Schütte, series***Proof Theory***, Springer, Berlin , 1970.

[30] Michael Sheard, *A guide to truth predicates in the modern era*, ***The Journal of Symbolic Logic***, vol. 59 (1994), pp. 1032–1054.

[31] Barry Hartley Slater, *Natural language's semantic closure*, forthcoming.

[32] Richmond H. Thomason, *A note on syntactical treatments of modality*, ***Synthese***, vol. 44 (1980), pp. 391–396.

[33] Albert Visser, *Semantics and the liar paradox*, ***Handbook of philosophical logic***, vol. IV, Reidel, Dordrecht, 1989, pp. 617–706.

UNIVERSITÄT KONSTANZ
FACHGRUPPE PHILOSOPHIE
POSTFACH 5560 D 21
78434 KONSTANZ, 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
GRADUATE SCHOOL OF SCIENCE & TECHNOLOGY
KOBE UNIVERSITY
ROKKO-DAI, NADA-KU
KOBE 657, JAPAN
Current address: Institut für formale Logik, Währinger Str. 25, 1090 Wien, Austria
E-mail: welch@kobe-u.ac.jp