Necessary Truths and Supervaluations

Carlo Nicolai
LMU Munich

Abstract

Starting with a trustworthy theory T, Galvan (1992) suggests to read off, from the usual hierarchy of theories determined by consistency strength, a finer-grained hierarchy in which theories higher up are capable of 'explaining', though not fully justifying, our commitment to theories lower down. One way to ascend Galvan's 'hierarchy of explanation' is to formalize soundness proofs: to this extent it often suffices to assume a full theory of truth for the theory T whose soundness is at stake. In this paper, we investigate the possibility of an extension of this method. Our ultimate goal will be to extend T not only with truth axioms, but with a combination of axioms for predicates for truth and necessity. We first consider two alternative strategies for providing possible worlds semantics for necessity as a predicate, one based on classical logic, the other on a supervaluationist interpretation of necessity. We will then formulate a deductive system of truth and necessity in classical logic that is sound with respect to the given (nonclassical) semantics.

1 HIERARCHIES OF THEORIES AND EVIDENCES

Logical complexity is one of the most fascinating and deep facts stemming from the incompleteness phenomena, and it is also one of the main themes of Sergio Galvan's ongoing journey into logic and philosophy. Just to mention a well-known example, the complexity of the set of elementary *truths* of a first-order theory[1] containing a modicum of arithmetic will always exceed – in a formally precise sense – the complexity of the set of *theorems* of that theory.

The mismatch between truth and provability is one of the central research interests of Sergio Galvan, as it became clear already with his first work on Tarski (Galvan 1973). The incompleteness theorems determine a hierarchy of 'natural' theories given by consistency strength or similar means of comparison. The consistency of Zermelo-Fraenkel set theory with choice ZFC can be proved for instance in ZFC

[1] Here by theory we always intend theory in classical logic.

plus the existence of the least wordly ordinal, which will then occupy a higher position than ZFC in the hierarchy. More generally, it is a consequence of Gödel's results that the consistency of a sufficiently rich T can only be proved in theories 'stronger' than T. Suitable set existence axioms, but also reflection principles, truth principles, have all been employed to properly extend T to theories that are capable of deriving its consistency or equivalent statements.

As Galvan (1992) lucidly points out, there is little epistemological interest in justifying the acceptance of T under assumptions stronger than T, at least if the kind of justification we are after is close to a fully-fledged *foundation*. Galvan's analysis of incompleteness, therefore, suggests to read-off, in the hierarchy of theories given by pure strength, a finer-grained hierarchy of *explanation*. To this hierarchy belong theories that are capable of formalizing and making explicit our commitment to theories lying lower down in the hierarchy. The theory PA+Con(PA) (cf. §2), obtained by adding a (intensionally correct) consistency statement to PA, will *not* belong to Galvan's hierarchy of explanation, although its consistency strength trivially exceeds the one of PA; the simple assumption of the consistency of PA does not represent in fact an explanation of our acceptance of PA, in Galvan's sense, but a mere stipulation. By contrast, the subsystem of second-order arithmetic ACA will belong to the hierarchy of explanation as it can define a full truth predicate for PA – more specifically, a full truth class for PA: this suffices for formalizing in ACA the metatheoretic proof of the soundness of PA.

It is therefore natural to assume that one way to climb up Galvan's hierarchy of explanation, given a trustworthy starting point T, is to assume a theory of truth for it. In this way one may achieve a sort of 'explanatory foundation' (Galvan 1992) – even though not a full justification of our trust in the base theory – rooted in our grasp of the notion of truth for T. There are several ways to add a theory of truth to a ground theory; a comprehensive treatment is Halbach (2014).

In this paper we investigate a possible extension of this method. One might see this work as an attempt to climb up Galvan's hierarchy of explanation by resorting to our grasp of 'logical' concepts such as truth itself but also of other modal notions, in primis necessity.[2] In other words we investigate the possibility of extending our base theory with 'natural' axioms governing modalities conceived as predicates and not as operators. This line of research is receiving new attention in the recent literature; Quine, Carnap, Montague have all already considered formal treatments of predicative uses of modal notions,[3] but the success of possible world semantics for operator modal logic and the presence of paradoxes in the predicate setting (see §3)

[2] We consider truth as a modal notion in the same vein of some medieval logicians such as William of Ockham. See for instance part II of the *Summa Logicae*.

[3] See for instance Carnap (1934), Quine (1960), Montague (1970).

have distracted much attention from it.

Halbach et alii (2003) have restored some confidence in the possibility of bridging the gap between modal logics and formal approaches to modal notions conceived as predicates. They have shown that, despite the presence of paradoxes, it is still possible to extend possible-worlds semantics to languages featuring modal predicates at least for some modal frames (cf. §4.1). Halbach&Welch (2009) have even suggested a generalization to arbitrary frames: a variant of their construction will be considered below.

The reader familiar with operator modal logics should not be worried: the predicate approach can be considered a generalization modal logics. Anything that can be said and proved in the operator approach can be mimicked in the predicate setting when suitable restrictions to the predicate language have been performed (see Gupta (1982) and Schweizer (1992)): it is in fact always possible to define an operator via a predicate. What we will say below will be no threat to the usual operator approach; paradoxes arise only when the expressive power of predicates and diagonalization comes into the picture. For the interested reader, Stern (2015) is a thorough and up to date treatment of the current research on syntactical treatments of modalities, including many original contributions by Stern himself.

We end this introductory section with three caveats. First of all we refer to truth, necessity, possibility, etc. as 'logical' notion in a rather liberal sense. Of course we do not advocate the view that the theories considered below amount to 'logics' in the very same sense in which first-order logic is 'logic'; rather we highlight the different possibilities that one faces when extending a given base theory. In this sense we oppose 'logical' principles, such as the ones characterizing *concepts* such as truth and necessity, to ontologically committing 'mathematical' principles, such as set existence assumptions. Furthermore, it is not our intention to suggest a *revision* of modal logics: the predicate approach, in our view, is a framework that naturally that naturally captures the ubiquitous predicative uses of modalities, and it is in this respect an interesting alternative to modal logics or its extensions. Finally, in this work we will only able to partially accomplish the promised ascent given by the combination of alethic modalities. Since the predicate approach to modalities is a lively but young field of research, there is some work required before tackling a fully-fledged proof-theoretic investigation of modal theories: in particular, as we shall see later on, consistency is a highly nontrivial matter.

Plan of The (rest of the) Paper. In §2 we introduce some of the preliminaries needed in the core sections of the paper. Further terminology and notation will be introduced in §4.1. In §3 we focus on some well-known paradoxes of the predicate approach such as Montague's, and on some less well-known antinomies essentially due to the interaction of more than one modal predicate. §4 will be devoted to possible worlds semantics for languages expanding our base language $\mathcal{L}$ with a primitive necessity predicate: we first describe some strategies available to retain a classical interpretation of necessity by restricting the modal space, and then remove these restrictions via a quasi-classical interpretation of necessity based on supervaluations. §5 will finally be devoted to deductive systems: we extend Cantini's theory of truth VF with axioms for necessity and prove its soundness with respect to a multimodal semantics obtained by adapting the semantics given in §4. We conclude in §6 with some comments to the content of the previous sections and sketch some possible extensions.

2 SOME PRELIMINARIES

Robinson's arithmetic Q is often considered to be the theoretical lower-bound for the derivability of non-intensional independence results such as Gödel's first incompleteness theorem, Tarski's and Montague's theorems. Let $\mathcal{L} = \{0, \mathsf{S}, +, \times\}$. The axioms of Q are the universal closures of the following formulas:

Q1	$\mathsf{S}x \neq 0$
Q2	$\mathsf{S}x = \mathsf{S}y \to x = y$
Q3	$x \neq 0 \to \exists y\,(x = Sy)$
Q4	$x + 0 = x$
Q5	$x + \mathsf{S}y = \mathsf{S}(x + y)$
Q6	$x \times 0 = 0$
Q7	$x \times \mathsf{S}y = (x \times y) + x$

The axiom Q3 is a weak form of induction and it indispensable to characterize the successor function, as in its absence there may be nonzero natural numbers without a predecessor. Q3 becomes derivable, however, when induction is added to Q.[4]

Peano arithmetic (PA) will play an important role in what follows: it is the result of adding to Q the schema of mathematical induction

$$\text{(Ind)} \qquad \varphi(0) \land \forall x(\varphi(x) \to \varphi(\mathsf{S}x)) \to \forall x\, \varphi(x)$$

for all $\mathcal{L}$-formulas $\varphi(v)$ with at most v free.

PA will be the theory formalizing the structure and properties of the bearers of modal ascriptions. We assume a standard arithmetization of the usual primitive recursive syntactic notions and operations of $\mathcal{L}$ and its extensions as it can be found, for instance, in Galvan (1992). In practice, we will work in a definitional extension of PA in which function symbols (e.g. for syntactic operations) for some primitive

[4] Q is extremely weak. Saul Kripke observed in fact that cardinal numbers are a model of Q, and thus there are entities, such as infinite cardinals, for which $\mathsf{S}x = x$.

recursive functions are available. They can however be eliminated in the usual way (see again Galvan (1992)).

As to notational conventions, we only give few instructive examples: $\underset{\cdot}{\neg} x$ stands for the $\mathcal{L}$-term representing in PA the operation of prefixing a negation symbol to x, and similarly $\underset{\cdot}{\mathsf{N}} t$ is the $\mathcal{L}$-term representing the result of prefixing the predicate N of the language $\mathcal{L}_{\mathsf{N}} := \mathcal{L} \cup \{\mathsf{N}\}$ to the object coded by t; $\mathsf{Sent}_{\mathcal{L}}(x)$ is a PA-definable formula representing the primitive recursive set of sentences of $\mathcal{L}$; the $\mathcal{L}$-formula $\mathsf{Bew}_T(x)$ represents the recursively enumerable set of theorems of the recursive theory T; x° stands for the PA-definable evaluation function assigning to each closed term its value. When this is clear from the context, we follow the customary practice and do not distinguish between sentences and their codes.

We conclude this section by introducing a technical device that is often useful to interpret self-applicable predicates. A sound translation function $\tau\colon \mathcal{L}_{\mathsf{N}} \to \mathcal{L}_1$ for sentences of the form $\mathsf{N}\underset{\cdot}{\mathsf{N}} t$, replacing $\mathsf{N}(\cdot)$ with some $\mathcal{L}_1$-formula $\xi(\cdot)$ should of course yield $\xi(\underset{\cdot}{\tau}\underset{\cdot}{\mathsf{N}} t)$ and not $\xi(\mathsf{N}t)$, where the notation $\underset{\cdot}{\tau}(\cdot)$ is like in the previous paragraph. To achieve the required translation, one may resort to the recursion theorem (Rogers 1987, §11.2), that yields for any recursive $f(x,y)$ an index e such that $f(e,y) = [e](y)$, where $[\cdot](\cdot)$ is the universal program. If we recursively define a function τ_0 such that, in the relevant case, $\tau_0(x, \mathsf{N}\mathsf{N}t) = \xi([x](\underset{\cdot}{\mathsf{N}} t))$, we would then be able to apply the recursion theorem and find an index e for τ_0 such that $[\mathsf{e}](\mathsf{N}\mathsf{N}t) = \xi([\mathsf{e}](\underset{\cdot}{\mathsf{N}} t))$. We are done by letting $\tau(x) \cong [\mathsf{e}](x)$.

3 Montague's paradox and extensions

Paradox is one of the main challenges that the proponent of the predicate approach to modalities has to face. In this section we introduce some paradoxical patterns of reasoning daunting the predicate approach by distinguishing the *unimodal* framework, in which our ground language is extended with only one modality, and a *multimodal* setting, in which more modalities are taken to interact. As it happens, paradox arises in both frameworks.

Montague's paradox is arguably the most fudamental form of paradoxicality in the unimodal setting. The theorem can be stated also in a more general form (Montague 1974), but here we shall be content with the following.

Lemma 1 (Montague)**.** *Let $T \supseteq \mathsf{Q}$ and assume there is a unary (possibly defined) predicate χ such that, for all $\varphi \in \mathcal{L}_T$:*

(T) $\qquad T \vdash \chi\ulcorner\varphi\urcorner \to \varphi$

(NEC) $\qquad$ *if* $T \vdash \varphi$, *then* $T \vdash \chi\ulcorner\varphi\urcorner$

Then T is inconsistent.

Proof. By the diagonal lemma, there is a sentence γ of $\mathcal{L}_T$ such that

$$T \vdash \gamma \leftrightarrow \neg\chi\ulcorner\gamma\urcorner$$

Now we reason in T as follows:

$\chi\ulcorner\gamma\urcorner \to \gamma$	(T)
$\chi\ulcorner\gamma\urcorner \to \neg\gamma$	def. γ
$\neg\chi\ulcorner\gamma\urcorner$	
γ	by def. γ
$\chi\ulcorner\gamma\urcorner$	(NEC)

qed

It is easy to see why Lemma 1 or variants thereof have led many authors, including Montague, to conclude that virtually no modal reasoning can be carried out in the predicate approach to modality. (T) and (NEC) are in fact basic for our understanding of some modalities, above all *de dicto* necessity.

This is, as we shall see shortly, a rather hasty conclusion. There are many examples of predicative uses of modalities in our philosophical reasoning, including core claims such as 'There are a posteriori necessary truths', or 'Any analytic judgment is necessary', that are most naturally formalized using modal predicates. Some portions of our reasoning with predicative modal asciptions can be rescued from paradox.[5]

One might argue at this stage that, as in the case of the Liar paradox, there is a straightforward way out of paradox given by Tarski's hierarchy of languages. If this is obviously true for the unimodal setting, when we move to languages featuring at least two modalities typing is not a sufficient solution anymore. Halbach (2006), for instance, produced the following, illuminating example involving two modalities M_1 and M_2 that closely resemble truth and necessity.

To formulate Halbach's result, let $T \supseteq \mathsf{Q}$ and expand $\mathcal{L}_T$ with predicates M_1 and M_2; call the resulting language $\mathcal{L}^+$. We say that $\varphi \in \mathcal{L}^+$ does not contain M_i if it does not contain any *used* occurrences of it, but it may contain *mentioned* occurrences.

Proposition 1 (Halbach). *Let T^+ extend T with the axiom schemata*

(1) $M_1\ulcorner\varphi\urcorner \leftrightarrow \varphi$ *for all $\varphi \in \mathcal{L}^+$ not containing M_1.*

(2) $M_2\ulcorner\varphi\urcorner \to \varphi$ *with $\varphi \in \mathcal{L}^+$ not containing M_2*

(3) $\dfrac{\varphi}{M_2\ulcorner\varphi\urcorner}$ *with $\varphi \in \mathcal{L}^+$ not containing M_2*

[5] Surely that are ways to strengthen the operator approach and mimic the expressive power of modal predicates, but one can hardly deny that the resulting formalizations will be less natural.

Then T^+ is inconsistent.

Proof. We reason in T^+:

$\nu \leftrightarrow \neg M_1 \ulcorner M_2 \ulcorner \nu \urcorner \urcorner$	diagonal lemma
$M_2 \ulcorner \nu \urcorner \rightarrow \neg \nu$	by (1)
$M_2 \ulcorner \nu \urcorner \rightarrow \nu$	(2)
$\neg M_2 \ulcorner \nu \urcorner$	
$\neg M_1 \ulcorner M_2 \ulcorner \nu \urcorner \urcorner$	by (1)
ν	def. ν
$M_2 \ulcorner \nu \urcorner$	(3)

qed

Our emphasis on the use\mention distinction should be now more motivated: the paradox would in fact disappear if instead of sentences of $\mathcal{L}^+$ we had chosen sentences of the ground language $\mathcal{L}_T$, where also mentioned occurrences of the modalities are not allowed.

Proposition 1 is only one of the paradoxes arising from the interaction of modal predicates. For instance, a paradox involving knowledge structurally similar to Proposition 1 can be found in Halbach (2008); Horsten & Leitgeb (2001) also show that seemingly innocuous assumptions on the structure of time lead to the inconsistency of the future. Proposition 1 was preferred to other choices for a simple reason: it suggests that multimodal paradoxes are somewhat harder to eradicate than their unimodal cousins.

Some authors have already set the basis for a systematic study of the multimodal paradoxes and their properties. A promising line of research consists for instance in applying insights from diagonal modal logics to analyse the structure of multimodal paradoxes. The fundamental idea behind this approach is to mimic the expressive power of arithmetic by considering propositional languages expanded with constants for modal ascriptions and a diagonal axiom for each of them. This boost in the expressive power provides enough information to analyse the 'logical' structure multimodal paradoxes. The interested reader may consult Egré (2005) and Fischer & Stern (2015) for further details.

4 MODELS FOR NECESSARY TRUTHS

What has been said in the last section strongly suggests extra care in handling expansions of $\mathcal{L}$ with modal predicates. Therefore in this section we start more humbly by providing a possible-worlds semantics for the expansion of $\mathcal{L}$ with a single necessity predicate N. In the next section we will see how to combine a truth predicate with the necessity predicate.

We exclusively focus on *de dicto* necessity, that is we consider only necessity ascriptions that apply to propositions, and not *de re* necessity ascriptions, which attribute a property to an (or possibly more) object by necessity. If the formalization of de dicto necessity as a unary predicate applying to names of sentences seems uncontroversial,[6] there are several options to deal with de re necessity or more generally de re modality. A promising option is to employ a binary predicate applying to unary formulas (playing the role of properties) and sequences of domain objects (variable assignments), mimicking a binary predicate for satisfaction. A careful treatment to de re modality, also in comparison to indexed modalities in modal logic, is deferred to a forthcoming work.

As we have mentioned in the introductory section, there are essentially to ways of constructing a possible world semantics for $\mathcal{L}_{\mathsf{N}} = \mathcal{L} \cup \{\mathsf{N}\}$. One can either consider a specific set of frames and allow for a classical interpretation of N, or instead impose no restrictions to the admissible frames and interpret the necessity predicate in a nonclassical way. We are mostly interested in the latter option, but for the sake of completeness we will also briefly sketch the fundamentals of the former without proofs: some terminology and the core insights of the classical approach will in fact also be useful later on.

4.1 Classical Interpretations of Necessity

We begin with some notions that may sound familiar from operator modal logic, but that it is worth repeating due to the new environment. They will also be useful in later sections.

Definition 1. *Models of $\mathcal{L}_{\mathsf{N}}$ will be pairs $(\mathbb{N}, X)$ where X is the extension of N. These pairs are 'worlds' in a possible worlds model. Therefore since we are dealing with standard models of $\mathcal{L}$ only, we may write (w, X) and $(\mathbb{N}, X)$ interchangeably.*

(i) A frame *is a pair (W, R) with $W \neq \varnothing$ and $R \subseteq W \times W$;*

(ii) A possible worlds model is a triple (W, R, V), with (W, R) a frame and V a function from worlds to subsets of $\mathcal{L}_{\mathsf{N}}$ such that for every $w \in W$:

$$V(w) = \{\varphi \in \mathcal{L}_{\mathsf{N}} \mid \forall u(wRu \Rightarrow V(u) \vDash \varphi)\}$$

[6] Obviously the controversy may arise at the level of the bearers of modal ascriptions. As usual, the sentence or the proposition are equally good candidates. Following the recent literature we take sentence types to be the bearers of modal ascritpions.

(iii) A frame (W, R) *admits a valuation if and only if there is a* V *such that* (W, R, V) *is a possible worlds model.*

We notice that, unsurprisingly, many basic consequences of the definitions carry over in the predicate approach. In particular, we have the standard properties of models of the operator modal logic K.

Lemma 2. *For* (W, R, V) *a possible worlds model and* $w \in W$*:*

(i) $(\mathbb{N}, V(w)) \models \mathsf{N}\ulcorner\varphi\urcorner$ *if and only if* $(\forall v \in W)(wRv \Rightarrow (\mathbb{N}, V(v)) \models \varphi)$

(ii) if $(\mathbb{N}, V(v)) \models \varphi$ *for all* $v \in W$, $(\mathbb{N}, V(w)) \models \mathsf{N}\ulcorner\varphi\urcorner$.

(iii) $(\mathbb{N}, V(w)) \models \mathsf{N}\ulcorner\varphi \to \psi\urcorner \wedge \mathsf{N}\ulcorner\varphi\urcorner \to \mathsf{N}\ulcorner\psi\urcorner$

Next we finally turn to the differences between the predicate and operator approach. If, given a frame (W, R) and worlds modelling $\mathcal{L}$, we can always construct a model for the language $\mathcal{L} \cup \{\Box\}$ by recursively defining truth for $\mathcal{L}_\Box$, the same strategy fails for the language $\mathcal{L}_\mathsf{N}$. Only certain frames admit a valuation, due to the paradoxical phenomena considered in the previous section. For instance, Lemma 1 shows that no reflexive frame admits a valuation for the necessity predicate.

One may wonder at this stage whether there are any general criteria to isolate the frames support a valuation. To this end, we introduce new terminology.

Definition 2.

(i) A (binary) relation R *is* converse well-founded *on a set* X *iff all nonempty* $Y \subseteq X$ *have an* R*-maximal element.*

(ii) A frame (W, R) *is converse well-founded iff* R *is converse well-founded on* W.

(iii) If (W, R) *is a frame and* R *converse well-founded on* W*, the* rank *of* $w \in W$ *is:*

$$\rho(w) := \begin{cases} 0, & \text{if there is no } v \text{ with } wRv \\ \alpha + 1, & \text{if } \exists v(wRv \wedge \rho(v) = \alpha \wedge \forall u(wRu \Rightarrow \rho(u) \leq \alpha)) \end{cases}$$

(iv) The converse well-founded *part of* $\{v \mid wRv\}$ *w.r.t.* W *is the largest* R*-upwards closed* $X \subseteq W$ *such that* R^{-1} *is well-founded on* X.

(v) The rank *of a converse ill-founded world is the rank of its converse well-founded part.*

If $\rho(w) = 0$, we say that w is a *dead end*.

Depending on the frames considered, it is possible to impose sufficient conditions on the existence of valuations. Converse well-foundedness is one of them. By transfinite induction on the rank of $w \in W$ in a converse well-founded frame (W, R) one defines the valuation

$$V(w) := \{\varphi \in \mathsf{Sent}_{\mathcal{L}_\mathsf{N}} \mid \forall v(wRv \Rightarrow (\mathbb{N}, V(v)) \models \varphi)\} \tag{4}$$

The crucial point is that if R is converse well-founded, we have the following picture for any $w \in W$,

In other words, from any w it is alway possible to reach a dead end, u_1 in this case, in finitely many steps. The valuation defined by (4) is thus unique, yielding

Proposition 2 (Gupta&Belnap)**.** *If (W, R) is converse well-founded, it admits a unique valuation.*

For frames containing converse ill-founded worlds, it is also possible to find a valuation under certain circumstances. To see this, let us consider the operator $\Phi\colon \mathcal{P}(\omega) \to \mathcal{P}(\omega)$

$$\Phi(X) := X \cap \{\varphi \in \mathsf{Sent}_{\mathcal{L}_\mathsf{N}} \mid (\mathbb{N}, X) \models \varphi\}$$

$\Phi(\cdot)$ is a decreasing and anti-monotone operator (i.e. $\Phi(Y) \subseteq Y$ for all Y and $\alpha \leq \beta$ entails that $\Phi^\alpha(Y) \supseteq \Phi^\beta(Y)$). Therefore, if one starts with $\Phi^0(\mathcal{L}_\mathsf{N}) := \Phi(\mathcal{L}_\mathsf{N})$ and iterates the application along an ordinal path – taking intersections at limit stages – one reaches a fixed point with associated a closure ordinal, that is a stage κ in which $\Phi^\kappa(\mathcal{L}_\mathsf{N}) = \Phi^\beta(\mathcal{L}_\mathsf{N})$ for all $\beta \geq \kappa$. The closure ordinal of $\Phi(\cdot)$ has also been computed by Halbach et alii (2003) as the least α such that the corresponding level of the constructible hierarchy L_α possesses a Σ_1-elementary end extension (Halbach et alii 2003, Prop. 21). In particular, we have $\kappa > \omega_1^{\mathsf{CK}}$, the first nonrecursive ordinal.

If a frame (W, R) is transitive and has converse ill-founded worlds w, the fixed point $\Phi^\kappa(\mathcal{L}_\mathsf{N})$ can always be used as valuation when the rank of w is greater than or equal to κ. That is

Proposition 3 (Halbach et alii (2003))**.** *If (W, R) is transitive and the rank of its converse ill-founded worlds is not smaller than κ, then (W, R) supports a valuation.*

The closure ordinal κ is also useful to impose necessary conditions on the existence of valuations in transitive frames. Let A be the class of admissible ordinals (without ω) with limits (see for instance Devlin (1984)).

Proposition 4 (Halbach et alii (2003)). *If* (W, R) *is transitive and admits a valuation, then for a converse ill-founded world* $w \in W$*, either* $\rho(w) \in \mathsf{A}$ *or* $\rho(w) \geq \kappa$.

Proposition 4 tells us that if (W, R, V) is a possible worlds model and R is converse ill-founded, then there will always be, for $w \in W$, an initial well-ordered portion of 'rank' $\alpha \in \mathsf{A}$ or greater-equal than κ. This means that frames (W, R) whose worlds have rank less than the first admissible ordinal ω_1^{CK} admit a valuation if and only if R is converse well-founded. Moreover, Proposition 4 can be generalized to non transitive frames, if we focus on the transitive closure of the accessibility relation.

In this brief overview our main intention was to highlight a fundamental fact: if one is interested in a classical interpretation of the necessity predicate, there are strong limitations one has to face. Again this is not a problem for the predicate approach if opposed to the operator approach, as we have already mentioned that the operator language can be straightforwardly translated in the predicate language. The problem is internal to the predicate approach. There is in fact an alternative to the classical approach sketched in this section: we can preserve the generality of the possible worlds semantics for operator modal logics if we move to a nonclassical setting.

4.2 Arbitrary Frames: Supervaluations

In this section we present a method for constructing possible worlds models for arbitrary frames (W, R). As before, worlds $w \in W$ are standard models of the ground language $\mathcal{L}$. The strategy is reminiscent of Kripke's fixed-point construction (Kripke 1975), which can be also seen as a method for generating models for $\mathcal{L}_{\mathsf{N}}$ in a reflexive frame $(\{w\}, R)$. To produce models for arbitrary W, one has to generalize Kripke's construction.

Halbach&Welch (2009) have proposed a similar generalization of Kripke's theory based on the Strong Kleene evaluation schema. We explore an alternative option and employ the supervaluational scheme introduced by Van Frassen (1966). We will highlight some nice features of supervaluations as opposed to the Strong Kleene approach after introducing few definitions and some of their consequences.

As before, $(w, F(w))$ will denote a model of $\mathcal{L}_{\mathsf{N}}$ in which w specifies the standard model of the ground language but X is now an evaluation function $F\colon W \to (\mathsf{Sent}_{\mathcal{L}_{\mathsf{N}}} \times \mathsf{Sent}_{\mathcal{L}_{\mathsf{N}}})$: at each world $w \in W$ it assigns *disjoint* extension and an antiextension to N. We also define an ordering $\leq$ between evaluation functions such that $F \leq G$ if, at any $w \in W$, $F(w)^+ \subseteq G(w)^+$ and $F(w)^- \subseteq G(w)^-$. We define a (binary) relation $\vDash_{\mathsf{vf0}}$ linking pairs (w, X) and $\mathcal{L}_{\mathsf{N}}$-sentences φ:[7]

$$(w, F(w)) \vDash_{\mathsf{vf0}} \varphi :\Leftrightarrow (\forall G)(F \leq G \wedge G(w)^{+} \subseteq \omega \setminus F(w)^{-} \Rightarrow (w, G(w)^{+}) \vDash \varphi)$$

The condition $G(w)^{+} \subseteq \omega \backslash F(w)^{-}$, together with the disjointness of extension and antiextension, will force consistent fixed points as extensions of the necessity predicate; that is, at any world w for no $\varphi \in \mathcal{L}_{\mathsf{N}}$, $\neg\varphi$ and φ will be in the ultimate extension of the necessity predicate. The relation $\vDash_{\mathsf{vf0}}$ extends the standard supervaluational picture according to which truth is satisfaction in all candidate extensions of a starting set; here we have merely generalized this picture to many worlds.[8]

To assign a suitable interpretation to the necessity predicate, we consider a variant of the strategy adopted by Halbach&Welch (2009) and impose further conditions on evaluation functions. We let EV be the set of such evaluations:

Definition 3. *The operator* $\Delta\colon \mathsf{EV} \to \mathsf{EV}$*, at each* $w \in W$*, is such that:*

$$(\Delta(F))(w)^{+} := \{\varphi \mid \forall v(wRv \Rightarrow (v, F(v)) \vDash_{\mathsf{vf0}} \varphi)\}$$
$$(\Delta(F))(w)^{-} := \{\varphi \mid \exists v(wRv \wedge (v, F(v)) \vDash_{\mathsf{vf0}} \neg\varphi)\}$$

The following is an immediate corollary of the definitions.

Corollary 1. *The operator* Δ *is monotone with respect to* $\leq$*, that is, for all* $w \in W$*,*

$$F \leq G \Rightarrow (\Delta(F))(w) \leq (\Delta(G))(w)$$

The monotonicity of Δ implies the existence of fixed points. This follows from abstract cardinality considerations (Moschovakis 1974, Thm. 1.A.1). As before, we may track the applications of Δ on an ordinal path using suitable indices. In other words $\Delta^{\alpha}(F)(w)$ denotes the α^{th} application of Δ to the starting evaluation function F at a world w, taking unions at limit stages. A fixed point of Δ will thus be an ordinal κ such that $\Delta^{\kappa}(F)(w) = \Delta^{\beta}(F)(w)$ for all $\beta \geq \kappa$.

By reflecting on the properties of $\Delta(\cdot)$, we have

Proposition 5. *If* F *is a fixed point of* Δ*, for all* $\varphi \in \mathcal{L}_{\mathsf{N}}$ *and frames* (W, R) *with* $w \in W$*:*

(5) $(w, F(w)) \vDash_{\mathsf{vf0}} \mathsf{N}\ulcorner\varphi\urcorner \;\Leftrightarrow\;$ *for all* v*, if* wRv*, then* $(v, F(v)) \vDash_{\mathsf{vf0}} \varphi$

(6) $(w, F(w)) \vDash_{\mathsf{vf0}} \neg\mathsf{N}\ulcorner\varphi\urcorner \;\Leftrightarrow\;$ *exists a* v *with* wRv *and* $(v, F(v)) \vDash_{\mathsf{vf0}} \neg\varphi$

[7] This is in a sense a simplifying choice: we dispense with variable assignments as we assume that we have constant domains and fixed names for all objects at every $w \in W$.

[8] There are other possible choices of the evaluational scheme, still in the supervaluational spirit. See Burgess (1986) or Fischer et alii (2015).

Proof.

Ad (5). ($\Rightarrow$) If $(w, F(w)) \vDash_{\mathsf{vf0}} \mathsf{N}\ulcorner\varphi\urcorner$, then for all evaluations $G \geq F$, including F itself: if $G(w)^+ \subseteq \omega \smallsetminus F(w)^-$, then $(w, G(w)^+) \vDash \mathsf{N}\ulcorner\varphi\urcorner$. Therefore, $\varphi \in F(w)^+$. Since F is a fixed point of Δ, $F(w)^+ = (\Delta(F))(w)^+$, and $\varphi \in (\Delta(F))(w)^+$, that is

$$\forall v(wRv \Rightarrow (v, F(v) \vDash_{\mathsf{vf0}} \varphi)$$

($\Leftarrow$) If for all v with wRv, $(v, F(v) \vDash_{\mathsf{vf0}} \varphi$, by definition of Δ also $\varphi \in (\Delta(F))(w)^+$. Again by the fixed point property, $\varphi \in F(w)^+$. This means that for all evaluations $G \,_1\!\geq F$, and a fortiori the ones in which $G(w)^+ \subseteq \omega \smallsetminus F(w)^-$, $\varphi \in G(w)^+$. By the classical satisfaction relation, $(w, G(w)^+) \vDash \mathsf{N}\ulcorner\varphi\urcorner$. By definition of $\vDash_{\mathsf{vf0}}$ we finally obtain

$$(w, F(w)) \vDash_{\mathsf{vf0}} \mathsf{N}\ulcorner\varphi\urcorner$$

Ad (6). ($\Rightarrow$): If $(w, F(w)) \vDash_{\mathsf{vf0}} \neg\mathsf{N}\ulcorner\varphi\urcorner$ the for all $G \geq F$ and $G(w)^+ \subseteq \omega \smallsetminus F(w)^-$, $\varphi \notin G(w)^+$. A fortiori, $\varphi \notin \omega \smallsetminus \mathsf{NSent}_{\mathcal{L}_\mathsf{N}} \smallsetminus F(w)^-$ where $\mathsf{NSent}_{\mathcal{L}_\mathsf{N}}$ is the set of numbers that are not $\mathcal{L}_\mathsf{N}$-sentences; therefore $\varphi \in F(w)^- = (\Delta(F))(w)^-$.

($\Leftarrow$). If $\exists v(wRv \wedge (v, F(v)) \vDash_{\mathsf{vf0}} \neg\varphi)$, then $\varphi \in (\Delta(F))(w)^- = F(w)^-$. Therefore $(w, G(w)^+) \vDash \neg\mathsf{N}\ulcorner\varphi\urcorner$ for any $G \geq F$ and $G(w)^+ \subseteq \omega \smallsetminus F(w)^-$, that is

$$(w, F(w)) \vDash_{\mathsf{vf0}} \neg\mathsf{N}\ulcorner\varphi\urcorner$$

qed

The *minimal* fixed point $\mathcal{I}_\Delta$ is obtained by closing the empty evaluation under Δ at any world $w \in W$, and it is the minimal fixed point that we now examine to highlight some nice features of the supervaluationist approach to necessity.

Let us call the *contingency teller* the sentence μ such that

$$\mathsf{Q} \vdash \mu \leftrightarrow \neg\mathsf{N}\ulcorner\mu\urcorner$$

Montague's paradox rules out reflexive frames in the classical setting. The contingency teller played an important role in the proof of Lemma 1. To see how the nonclassical setting helps in dealing with paradoxes, we now show that in the new setting μ will be 'gappy', that is neither necessary or contingent, in the minimal fixed point.

Lemma 3. *Let (W, R) be a frame. The contingency teller is neither in $\mathcal{I}^+_\Delta(w)$ nor in $\mathcal{I}^-_\Delta(w)$ for any $w \in W$.*

Proof. We prove the claim by induction on the construction of the minimal fixed point of Δ.

At stage $\Delta^0(\varnothing)(w) := (\varnothing, \varnothing)$, the claim is trivially satisfied.

At successor stages $\alpha + 1$, if $\mu \in \mathcal{I}_\Delta^{\alpha+1}(w)^+$, then

$$\forall v(wRv \Rightarrow (v, \mathcal{I}_\Delta^\alpha(v)) \vDash_{\mathsf{vf0}} \neg \mathsf{N}\ulcorner\mu\urcorner) \tag{7}$$

That is, $\mu \notin G(v)^+ \subseteq \omega \smallsetminus \mathcal{I}_\Delta^\alpha(v)^-$ for all suitable G, including $\omega \smallsetminus \mathsf{NSent}_{\mathcal{L}_\mathsf{N}} \smallsetminus \mathcal{I}_\Delta^\alpha(v)^-$. Therefore $\varphi \in \mathcal{I}_\Delta^\alpha(v)^-$, quod non by induction hypothesis.

If, by contrast, $\mu \in I_\Delta^{\alpha+1}(w)^-$, there will be, for all extensions G of $\mathcal{I}_\Delta^\alpha$, $(v, G(v)^+) \vDash \mathsf{N}\ulcorner\mu\urcorner$ at some accessible v. Thus $\mu \in I_\Delta^\alpha(v)^+$, again contradicting the induction hypothesis.

Finally, if $\mu \in \mathcal{I}_\Delta^\lambda$ for a limit λ, the claim follows from the previous steps by definition of $\Delta(\cdot)$. qed

By suitably adapting the argument of Lemma 3, one easily shows that $\neg\mu$ cannot be in $\mathcal{I}_\Delta$. Moreover, a generalization of this arguments shows that there are *consistent* fixed points of Δ.

As we have already observed, the operator $\Delta(\cdot)$ compares to the operator based on the Strong Kleene evaluation schema considered in Halbach&Welch (2009). It is well-known since Kripke (1975) that the Strong Kleene schema yields an attractive picture of self-applicable truth predicate. Above all, it yields a compositional semantics, e.g. $A \vee B$ is true$_{\mathsf{sk}}$ if and only if A is true$_{\mathsf{sk}}$ or B is true$_{\mathsf{sk}}$ with A, B sentences of a base language such as $\mathcal{L}$ plus a primitive truth predicate.

If necessity and not truth simpliciter is at stake, one may argue that compositionality is not as important as, for instance, establishing the necessity of all laws of classical logic; so $A \vee \neg A$ should be necessary even though we do not have the resources to find out whether A or its negation are true. The following results show that the supervaluationist approach captures, in the predicate approach, the picture of necessity just sketched.

Proposition 6.

(i) All logical laws, including the laws of the conditional (e.g. $\varphi \to \varphi$ for $\varphi \in \mathcal{L}_\mathsf{N}$) valid in $\mathcal{I}_\Delta$ (i.e. in $\mathcal{I}_\Delta^+(w)$ at any w);

(ii) Let PAN *simply* PA *formulated in $\mathcal{L}_\mathsf{N}$. All theorems of* PAN *are valid in $\mathcal{I}_\Delta$.*

Proof. In both cases one reflects on the definition of $\Delta(\cdot)$. At stage 1 of the construction of $\mathcal{I}_\Delta(w)$ for arbitrary w, we have

$$(\Delta^1(\varnothing,\varnothing))(w) = \left\langle \begin{array}{c} \{\varphi \mid \forall v(wRv \Rightarrow (v,(\varnothing,\varnothing)) \vDash_{\mathsf{vf0}} \varphi)\}, \\ \{\varphi \mid \exists v(wRv \& (v,(\varnothing,\varnothing)) \vDash_{\mathsf{vf0}} \neg\varphi)\} \end{array} \right\rangle$$

By definition of $\vDash_{\mathsf{vf0}}$, therefore, all theorems of first-order logic and of PAN will get in $(\Delta^1(\varnothing,\varnothing))(w)^+$; therefore by the monotonicity of Δ also in $\mathcal{I}_\Delta(w)^+$.

qed

As a corollary, $\mathsf{N}\ulcorner\mu\vee\neg\mu\urcorner$ will be valid in the fixed point $\mathcal{I}_\Delta$ at any world, although μ, as we have seen already, will not be in any fixed point. In addition, also tricky biconditionals such as $\mu \leftrightarrow \neg\mathsf{N}\ulcorner\mu\urcorner$ will be in the fixed point.

We have thus seen that there are ways to overcome the paradoxes of the predicate approach and capture predicative uses of necessity by providing *models* for the base language expanded with a predicate for necessity. In the next section we consider some strategies to formulate deductive systems inspired to the semantic construction just given.

5 A SYSTEM FOR TRUTH AND NECESSITY

In this section we move the first steps into combining truth and necessity. We introduce a modal version of Cantini's VF (Cantini 1990) and prove its soundness with respect to a modification of the semantics given in the previous section.

5.1 The theory VF

VF is the theory capturing the properties of a self-applicable (type-free) truth predicate interpreted according to a suitable modification of the operator Δ introduced above:

$$(\mathbb{N}, X) \vDash_{\mathsf{vf}} \varphi :\Leftrightarrow \forall S(X \subseteq S \wedge \mathsf{con}^*(S) \Rightarrow (\mathbb{N}, S) \vDash \varphi$$

Here we have dropped the antiextension and we deal only with consistent candidate extension: in particular $\mathsf{con}^*(G(w))$ expresses that $G(w)$ does not contain negations of sentences in $F(w)$; to avoid triviality, *only consistent* starting evaluations $F(w)$ are allowed. Let $\mathcal{L}_\mathsf{T}$ be the language $\mathcal{L}$ expanded with a unary truth predicate T. We call the new operator $\Theta\colon\mathcal{P}(\omega)\to\mathcal{P}(\omega)$:

$$\Theta(X) := \{\varphi \mid (\mathbb{N}, X) \vDash_{\mathsf{vf}} \varphi\}$$

By only a slight modifications of the arguments already given there, we notice that Θ is monotone and thus it has fixed points. We define by transfinite induction

$$\mathcal{I}^0_\Theta := \varnothing$$
$$\mathcal{I}^{\alpha+1}_\Theta := \Theta(\mathcal{I}^\alpha_\Theta)$$
$$\mathcal{I}^\lambda_\Theta := \bigcup_{\beta<\lambda} \mathcal{I}^\beta_\Theta$$

The minimal fixed point $\mathcal{I}_\Theta$ is simply $\mathcal{I}^\kappa_\Theta$, where κ is the closure ordinal for Θ.

Cantini (1990) introduced a deductive system that is sound with respect to fixed points of Θ. It is called VF from 'Van Frassen', who first introduced the supervaluational scheme to analyse vague predicates.

Definition 4. VF *is formulated in* $\mathcal{L}_{\mathsf{T}}$. *Its axioms are all axioms of* PAT *(i.e.* PA *formulated in* $\mathcal{L}_{\mathsf{T}}$*) and the following:*

(VF1)	$\forall \vec{x}\big(\mathsf{T}\ulcorner\varphi(\dot{\vec{x}})\urcorner \to \varphi(\vec{x})\big)$ *for all* $\varphi \in \mathcal{L}_{\mathsf{T}}$
(VF2)	$\forall s,t\big((\mathsf{T}(s \underset{\cdot}{=} t) \leftrightarrow s^\circ = t^\circ) \wedge (\mathsf{T}(s \underset{\cdot}{\neq} t) \leftrightarrow s^\circ \neq t^\circ)\big)$
(VF3)	$\forall x\big(\mathsf{Ax}_{\mathsf{PAT}}(x) \to \mathsf{T}x\big)$
(VF4)	$\forall v \forall x \forall t\,\big(\mathsf{T}\,x(t/v) \to \mathsf{T}\underset{\cdot}{\forall} v x\big)$
(VF5)	$\forall t\big(\mathsf{T}t^\circ \to \mathsf{T}\underset{\cdot}{\mathsf{T}}t\big)$
(VF6)	$\forall s\big(\mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(s^\circ) \wedge \mathsf{T}\underset{\cdot}{\neg}\mathsf{T}x \to \mathsf{T}\underset{\cdot}{\neg}s^\circ\big)$
(VF7)	$\forall x,y\big(\mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x \underset{\cdot}{\to} y) \to (\mathsf{T}(x \underset{\cdot}{\to} y) \to \mathsf{T}x \to \mathsf{T}y)\big)$
(VF8)	$\forall x(\mathsf{T}\ulcorner\mathsf{T}\dot{x} \to \neg\mathsf{T}\underset{\cdot}{\neg}\dot{x}\urcorner)$
(VF9)	$\mathsf{T}\ulcorner\mathsf{T}\dot{x} \to \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(\dot{x})\urcorner$

It is a routine task to check, by induction on the length of the derivation in VF, that

Proposition 7 (Cantini (1990), Prop. 3.4)**.** *If X is a fixed point of Θ, then* $(\mathbb{N}, X) \vDash \mathsf{VF}$.

5.2 Modal extensions of VF

To introduce a modal extension of VF, we consider a variant of the strategy adopted by Stern (2014) to extend the Kripke-Feferman system KF.[9]

We first introduce predicative counterparts of the well-known modal principles (T), (4) and (E) formulated in the language $\mathcal{L}_{\mathsf{TN}} := \mathcal{L} \cup \{\mathsf{T}\} \cup \{\mathsf{N}\}$:

(T)	$\forall x(\mathsf{Sent}_{\mathcal{L}_{\mathsf{TN}}} \wedge \mathsf{N}x \to \mathsf{T}x)$
(4)	$\forall t\,(\mathsf{T}\underset{\cdot}{\mathsf{N}}t \to \mathsf{N}\underset{\cdot}{\mathsf{N}}t)$
(E)	$\forall t\,(\mathsf{T}\underset{\cdot}{\neg}\underset{\cdot}{\mathsf{N}}t \to \mathsf{N}\underset{\cdot}{\neg}\underset{\cdot}{\mathsf{N}}t)$

As it is well-known from operator modal logic, (T) forces reflexive frames, (4) transitive frames, and (E) Euclidean frames. (T) in combination with (E) suffice to force frames based on an equivalence relation.

We finally define the theory MVF. The theory PATN is, as one might expect, simply PA formulated in $\mathcal{L}_{\mathsf{TN}}$.

[9] See again Halbach (2014) for a thorough introduction to KF.

Definition 5 (Modal VF). MVF *is the theory in* $\mathcal{L}_{\mathsf{TN}}$ *whose axioms are (i) the axioms of* PATN, *(ii)* VF *formulated in* $\mathcal{L}_{\mathsf{TN}}$, *(iii) the following sentences and rules:*

(T-in) $\forall t\,(\mathsf{N}t^\circ \to \mathsf{T}\underset{\cdot}{\mathsf{N}}t)$

(BF) $\forall v\forall x\big(\mathsf{Sent}_{\mathcal{L}_{\mathsf{TN}}}(\underset{\cdot}{\forall}vx) \to (\forall t\,\mathsf{N}x(t/v) \to \mathsf{N}\,\underset{\cdot}{\forall}vx)\big)$

(Rig1) $\forall s,t\forall v\forall x\,(\mathsf{Sent}_{\mathcal{L}_{\mathsf{TN}}}(\underset{\cdot}{\forall}vx) \to (s^\circ = t^\circ \to (\mathsf{N}x(s/v) \leftrightarrow \mathsf{N}x(t/v))))$

(Rig2) $\forall s\forall t(s^\circ \neq t^\circ \to \mathsf{N}(s \underset{\cdot}{\neq} t))$

(K) $\forall x\forall y\big(\mathsf{Sent}_{\mathcal{L}_{\mathsf{TN}}}(x \underset{\cdot}{\to} y) \to (\mathsf{N}(x \underset{\cdot}{\to} y) \to (\mathsf{N}x \to \mathsf{N}y))\big)$

$$(\mathsf{Nec})\ \frac{\mathsf{T}\ulcorner\varphi\urcorner}{\mathsf{N}\ulcorner\varphi\urcorner} \qquad \textit{for all } \varphi \in \mathcal{L}_{\mathsf{TN}}$$

It is worth emphasising that the axiom VF3 declaring the truth of all axioms of PAT now becomes

$$\forall x(\mathsf{Ax}_{\mathsf{PATN}}(x) \to \mathsf{T}x) \tag{8}$$

As before, by a straightforward induction, we can conclude that all theorems of PATN are true. This includes, for instance, all instances of excluded middle in the language $\mathcal{L}_{\mathsf{TN}}$.

As we have seen in the case of the paradoxes of interaction, eradicating inconsistencies in the multimodal framework is more difficult than in the unimodal setting. Therefore we first ensure that MVF is consistent by reducing its consistency to the consistency of VF. This will also give an upper bound to the proof-theoretic strength of MVF that will be discussed further in the concluding section. The lower bound is clear as VF is contained in MVF.

Proposition 8. MFV *is consistent, if* VF *is.*

Proof. We define the primitive recursive translation $\tau\colon \mathcal{L}_{\mathsf{TN}} \to \mathcal{L}_{\mathsf{T}}$ as follows, using the remarks at the end of §2:

$\tau(\varphi) := \varphi$ for $\varphi \in \mathcal{L}$ that is, φ arithmetical
$\tau(\mathsf{T}\ulcorner\varphi\urcorner) := \mathsf{T}\underset{\cdot}{\tau}\ulcorner\varphi\urcorner$
$\tau(\mathsf{N}\ulcorner\varphi\urcorner) := \mathsf{T}\underset{\cdot}{\tau}\ulcorner\varphi\urcorner$
τ commutes with propositional connectives and quantifiers

In essence, the translation just maps necessity into truth. It is easy to verify that the translations of all axioms of MVF are provable in VF. qed

Cantini (1990) showed that VF proves the same arithmetical sentences as the theory ID_1 of elementary positive inductive definitions (see Pohlers (2009)). Proposition 8, therefore, yields the following analysis of MVF.

Corollary 2. MVF *proves the same arithmetical sentences as* ID_1.

5.3 Semantics and Soundness

Proposition 8 gives us a consistency proof for MVF and indirectly a semantics for it; in any model of VF we can construct an internal model of MVF. This does not mean, however, that there are 'nice' models of MVF: in this section we show that there are 'standard models' of MVF obtained by generalizing in a rather natural way the intended models of VF.

We now adapt the semantics given in §4.2 to the multimodal framework. Given a frame $\mathcal{F}$, a model of the language $\mathcal{L}_{\mathsf{TN}}$ at a world $w \in W$ (again we think of $w \in W$ as standard models of $\mathcal{L}$) will be a triple $\mathcal{M}_w := (w, E(w), N_{E(w)})$, where $E\colon W \to \mathcal{P}(\omega)$ is a function assigning to each w an extension of the truth predicate. From this extension one standardly defines an extension of the necessity predicate $N_{E(w)}$ by taking the intersection of the set of truths at all accessible worlds:

$$N_{E(w)} := \{\varphi \in \mathcal{L}_{\mathsf{TN}} \mid \forall v(wRv \Rightarrow \varphi \in E(v))\}$$

The set of truths at all accessible worlds will then be defined using again the supervaluational scheme, but this time to define the extension of the truth predicate and not of the necessity predicate directly. Notice now that we can drop the superscript $^+$ or $^-$ as we are only assigning candidate extensions and not also an antiextension to the predicate. As before, let $\leq_1$ an ordering of the evaluation functions defined by: $E_0 \leq_1 E_1$ if and only if for all $w \in W$, $E_0(w) \subseteq E_1(w)$. Therefore we set, for $\varphi \in \mathcal{L}_{\mathsf{TN}}$:

$$\begin{aligned}&(w, F(w), N_{F(w)}) \vDash_{\mathsf{vf1}} \varphi :\Leftrightarrow \\ &\qquad (\forall G\ {}_1\!\geq F)(\mathsf{con}^*(G(w)) \Rightarrow (w, G(w), N_{G(w)}) \vDash \varphi)\end{aligned}$$

with $N_{X(w)}$ as above. With $\mathsf{con}^*(G(w))$ we mean again that $G(w)$ does not contain negations of sentences in $F(w)$; as above, to avoid triviality, we consider *only consistent* starting evaluations $F(w)$. The operator $H^{\mathcal{F}}$ on evaluation functions, relative to a frame $\mathcal{F}$, is defined as

$$(H^{\mathcal{F}}(E))(w) := \{\varphi \in \mathcal{L}_{\mathsf{TN}} \mid (w, E(w), N_{E(w)}) \vDash_{\mathsf{vf1}} \varphi\}$$

The following is an immediate consequence of the definitions.

Lemma 4. *The operator $H^{\mathcal{F}}$ is monotone with respect to $\preceq_1$.*

As before, monotonicity implies the existence of fixed points, that is evaluations such that $H^{\mathcal{F}}(E) = E$. In a fixed point of $H^{\mathcal{F}}(E)$, therefore, for any $\varphi \in \mathcal{L}_{\mathsf{TN}}$, and any world in $\mathcal{F}$,

(9)
$$(w, E(w), N_{E(w)}) \vDash_{\mathsf{vf1}} \mathsf{T}\ulcorner\varphi\urcorner \Leftrightarrow (\forall G \;{}_1\!\succeq E)(\mathsf{con}^*(G(w)) \Rightarrow (w, G(w), N_{G(w)}) \vDash \varphi)$$

(10)
$$(w, E(w), N_{E(w)}) \vDash_{\mathsf{vf1}} \mathsf{N}\ulcorner\varphi\urcorner \Leftrightarrow \forall v(wRv \Rightarrow \varphi \in E(v))$$

Closing the empty evaluation function under iterated applications of $H^{\mathcal{F}}$ along an ordinal path, we reach the minimal fixed point of $\mathcal{I}_{H^{\mathcal{F}}}$ of $H^{\mathcal{F}}$. MVF, however, is not only sound with respect to the minimal fixed point, but it is satisfied by *all* fixed points of $H^{\mathcal{F}}$.

Proposition 9. *Let $\mathcal{F} = (W, R)$ be a frame and R an equivalence relation. If $H^{\mathcal{F}}(E) = E$, then $(w, E(w), N_{E(w)}) \vDash \mathsf{MVF}$ for any $w \in W$.*

Proof. For the PATN axioms and the axioms of VF one merely adapts Cantini's proof. We consider the genuinely modal axioms of MVF.

Ad (T). If $(w, E(w), N_{E(w)}) \vDash \mathsf{N}\ulcorner\varphi\urcorner$, then $\varphi \in E(v)$ for all v accessible from w. By reflexivity, $\varphi \in E(w)$.

Ad ($\mathsf{4}$). Without loss of generality, we can reason about (the code of) a sentence φ. Let us assume $\mathsf{N}\ulcorner\varphi\urcorner \in E(w)$. This entails:

(11) $$(\forall G \;{}_1\!\succeq E)(\mathsf{con}^*(G(w)) \Rightarrow (w, G(w), N_{G(w)}) \vDash \mathsf{N}\ulcorner\varphi\urcorner)$$

Therefore, for all extended evaluations G, $\varphi \in N_{G(w)}$, that is

(12) $$\forall v(wRv \Rightarrow \varphi \in G(v))$$

Now assume $(w, E(w), N_{E(w)}) \vDash \neg\mathsf{N}\underset{\cdot}{\mathsf{N}}\ulcorner\varphi\urcorner$. There is then a v with wRv and $\mathsf{N}\ulcorner\varphi\urcorner \notin E(v)$, that is

(13) $$\exists v(wRv \wedge (\exists G \;{}_1\!\succeq E)(\mathsf{con}^*(G(v)) \wedge (v, G(v), N_{G(v)}) \vDash \neg\mathsf{N}\ulcorner\varphi\urcorner)$$

By iterating the same reasoning for $N_{G(v)}$, we find

(14) $$\exists v_0(vRv_0 \wedge \varphi \notin G(v_0))$$

By transitivity, (14) contradicts (12).

The reasoning for (E) is similar to the previous case. So we consider (T-in). If $t^\circ \in N_{E(w)}$, then $t^\circ \in E(v)$ for all v accessible from w. If $\mathsf{N}t \notin E(w)$, then

$$(\exists G \,{}_1\!\geq E)(\mathsf{con}^*(G(w)) \wedge (w, G(w), N_{G(w)}) \not\models \mathsf{N}t) \tag{15}$$

Again we arrive at

$$\exists v(wRv \wedge t^\circ \notin G(v)) \tag{16}$$

But (16) contradicts $G \,{}_1\!\geq E$ and $t^\circ \in E(v)$.

We conclude the proof by considering the case of (K). Let us assume $\varphi \to \psi \in N_{E(w)}$ and $\varphi \in N_{E(w)}$; that is

$$\forall v(wRv \Rightarrow \varphi \to \psi \in E(v) \wedge \varphi \in E(v)) \tag{17}$$

The definition of classical satisfaction yields the desired result. qed

In the following, concluding section we elaborate on the results just presented and on the possibility of further work.

6 CONCLUSION

In the introduction we sketched a project: the formulation of natural systems of interacting modalities extending a some trustworthy theory of modal ascriptions. The 'naturality' criterion imposed on the project has been spelled out in terms of a possible worlds semantics for modal predicates. We have seen that this is a nontrivial matter; paradoxes threaten our predicative uses of modal notions and impose severe restrictions to the space of models of the corresponding languages.

Despite these difficulties, we formulated a system of truth and necessity MVF that adapts the motivation behind Cantini's VF to the new language and that is sound with respect to a rather natural semantics for truth and necessity. In §1 we have recalled Sergio Galvan's idea of a hierarchy of theories that are capable of making explicit our trust in the theories lower down, starting with our preferred theory of the bearers of modal ascriptions (e.g. PATN). By Proposition 8, MVF will prove the same arithmetical sentences as ID_1: to give an idea of how this relates to what we called Galvan's hierarchy, we notice that, for instance, the results of iterating ACA for all ordinals $< \Gamma_0$, the so-called Feferman-Schütte ordinal, is reducible to ID_1. Reflecting on the fact that ACA was already sufficient to formalize the metatheoretic soundness proof for PA, this gives us an idea of how far MVF takes us.

There are however, also some limitations to the success of the strategy of combining modalities. In the first place it does not seem to be possible to achieve a full adequacy result for MVF, exactly as in the case of VF. More precisely, Fischer et alii (2015) have proposed the following criterion of adequacy for systems of truth: a system T is ω-categorical if

$$(\mathbb{N}, S) \models T \Leftrightarrow S \in \mathfrak{M} \tag{18}$$

where $\mathfrak{M}$ is a class of acceptable interpretations of the truth predicate given by some semantic theory of truth. Proposition 7 tells us that the right-to-left direction holds for VF. Fischer et alii (2015), by adapting a previous result of Philip Welch, show in fact that the left-to-right direction of (18) cannot be achieved if $\mathfrak{M}$ is the class of supervaluational fixed-points: in nuce, the property of being a supervaluational fixed point is Π^1_1-complete, if (18) held, we would have a Σ^1_1-definition of a supervaluational fixed point. This shows also that such categoricity result cannot be achieved for MVF.

In addition, we know already that Proposition 8 shows that the proof-theoretic strength of MVF does not exceed the one of VF. This is in some sense good news; at the conceptual level we might even welcome the fact that the notion of necessity axiomatized by MVF is in continuity with the corresponding notion of truth and allows for a 'collapse' of necessity into truth in the one world reading. However, it is also true that the interaction of truth and necessity enriches our expressive capability and we would like our modal theory to exceed the strength of the truth theory on which it is based.

These drawbacks of the strategy proposed in this paper may be nonetheless good guiding principles for adopting different strategies to combine truth and necessity: one might for instance follow McGee (1991) and Halbach (2001) and consider necessity as provability in a suitable system. We defer a treatment of this option to further work, but it will most likely lead to a considerable increase in proof-theoretic strength.

References

Burgess, J. (1987). The truth is never simple. *Journal of Symbolic Logic* 51:663–81.

Cantini, A. (1990). A theory of formal truth arithmetically equivalent to ID_1. *The Journal of Symbolic Logic* 55: 1.

Carnap, R. (1934). *Logische Syntax der Sprache*. Springer, Wien.

Devlin, K. (1984). *Constructibility*. Springer.

Egré, P. (2005). The knower paradox in the light of provability interpretations of modal logic. *Journal of Logic, Language and Information* 14: 13–48.

Fischer, M., V. Halbach, J. Krin, J. Stern (2015). Axiomatizing semantic theories of truth?. *The Review of Symbolic Logic*.

Fischer, M. and J. Stern (2015). Paradoxes of Interaction?. *Journal of Philosophical Logic* 44(3), 287–308.

Galvan, S. (1973). Il concetto di verità in A. Tarski. *Verifiche* 2: 3–66.

Galvan, S. (1992). *Introduzione ai teoremi di incompletezza*. Franco Angeli.

Galvan, S. (2006). *Einführung in die Unvollständigkeit Theoreme*. Padenborn.

Gupta, A. (1982). Truth and Paradox. *Journal of Philosophical Logic*.

Halbach, V. (2001). Disquotational truth and analyticity. *Journal of Symbolic Logic*.

Halbach, V. (2006). How not to state the T-sentences. *Analysis* 66, 276-280.

Halbach, V. (2008). On a side effect of solving Fitch's paradox by typing knowledge. *Analysis* 68, 114–120.

Halbach, V. (2014). *Axiomatic Theories of Truth*. Cambridge University Press.

Halbach V., H. Leitgeb and P. Welch (2003). Possible world semantics for modal notions conceived as predicates. *Journal of Philosophical Logic* 32: 179–233.

Halbach, V. and P. Welch (2009). Necessities and necessary truths: a prolegomenon to the use of modal logic in the analysis of intensional notions. *Mind* 118: 71–100.

Horsten, L. and Leitgeb, H. (2001). No Future. *Journal of Philosophical Logic* 30(3): 259–265.

McGee, V. (1991). *Truth, Vagueness and Paradox. An essay on the logic of truth.* Hackett.

Montague, R. (1970). Universal Grammar. *Theoria* xxxvi: 373–98.

Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy*.

Montague, R. (1974). Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability. In R. Thomason (ed.). *Formal Philosophy. Selected Papers of Richard Montague.* Yale University Press 1974.

Moschovakis, Y. (1974). *Elementary Induction on Abstract Structures*. North Holland.

Quine, W.V.O. (1960). *Word and Object*. MIT Press.

Pohlers, W. (2009). *Proof Theory. First Steps into Impredicativity.* Springer.

Rogers, H. (1987). *Theory of Recursive Functions and Effective Computability.* MIT Press.

Stern, J. (2015). *Towards a Predicate Approach to Modal Notions*. Springer.

Stern, J. (2014). Modality and Axiomatic Theories of Truth II: Kripke-Feferman. *The Review of Symbolic Logic* 7(2): 299–318.

Schweizer, P. (2002). Asyntactical apporach to modality. *Journal of Philosophical*

Logic 24: 451–454.

Tait, W. (1981). Finitism. *Journal of Philosophy*.

Van Frassen, B. (1966). Singular Terms, Truth-value Gaps and Free Logic. *Journal of Philosophy* 63: 481–495.