Paradoxes of Interaction?*

Johannes Stern†& Martin Fischer†
Munich Center for Mathematical Philosophy, LMU Munich

May 28, 2014

Abstract

Since Montague's work it is well known that treating a single modality as a predicate may lead to paradox. In their paper "No Future", Horsten and Leitgeb [8] show that if the two temporal modalities are treated as predicates paradox might arise as well. In our paper we investigate whether paradoxes of multiple modalities, such as the No Future paradox, are genuinely new paradoxes or whether they "reduce" to the paradoxes of single modalities. In order to address this question we develop a notion of reducibility based on a version of Smoryński Diagonalized Operator Logic. We show that there are reducible multimodal paradoxes as well as irreducible paradoxes of interaction. In particular, we show the No Future paradox to be an irreducible paradox according to our notion of reducibility.

1 Introduction

In philosophical logic modalities are commonly treated as operators. Yet, they may also be treated as predicates of sentences or propositions, an approach which was advocated by, e.g., Carnap, Quine or more recently Halbach et al. [7]. It is well known that predicate approaches to modality have to face liar like paradoxes such as, for example, the Knower paradox. These paradoxes of a single modality have been known since the 60's due to the work of Montague [10], Kaplan and Montague [9], Myhill [11] and others. More recently, examples of paradoxes involving multiple modalities have been discussed in the literature. One example is given by Horsten and Leitgeb [8]. They show that if the two temporal modalities of standard temporal logic are conceived as predicates, then paradox arises when

*To appear in the *Journal of Philosophical Logic*.

†Munich Center for Mathematical Philosophy; LMU Munich; Fakultät für Philosophie, Wissenschaftstheorie und Religionswissenschaft; Geschwister-Scholl-Platz 1; 80539 München; Germany. Email Johannes Stern: johannes.stern@lmu.de. Email Martin Fischer: m.fischer1@lmu.de.

very basic assumptions on behalf of the temporal modalities are made. Another example is given by Halbach [6]. Halbach shows that the principle of knowability which connects the notion of knowledge and an alethic modality produces inconsistency if knowledge is assumed to be factive and the alethic modality is closed under the rule of necessitation.

Although paradoxes of multiple modalities are not surprising per se—if we encounter paradox when dealing with a single modality, then there will be paradox in the case of multiple modalities likewise—the two mentioned examples seem to introduce a new aspect into the picture: it appears to be the interplay of the two modalities that creates the paradox, as opposed to the assumptions made on behalf of the two individual *modal notions*. That is, whereas in the case of the unimodal paradoxes the *modal properties* we assumed for the modal notion lead to paradox, these modal properties do not seem to be the cause of the paradoxes in the two examples. It should be possible to consistently adopt these modal properties. Rather, paradox seems to arise because of the assumptions we made on the interaction of the different modal notions, that is on the assumptions we made on how the two modal notions relate. However, in the absence of a clear criterion of what it means for a paradox to arise from the interaction of two (or more) modal notions the former assessment seems to be nothing more than a hunch and it remains unclear whether the paradoxes of multiple modalities contribute a new and interesting aspect to the discussion.

In this paper we investigate this question. In particular, we shall distinguish between two types of paradoxes of interaction: the paradoxes which arise due to the modal properties assumed for the individual modal notion and the proper paradoxes of interaction which arise due to the interaction of the different modal notions. The former paradoxes will be called reducible to the unimodal paradoxes, the latter are called irreducible. To this end we develop an appropriate notion of reducibility which will be spelled out using Smoryński's Diagonal Modal Logic (cf. Smoryński [14]). We shall then put the notion of reducibility to work and show that Horsten and Leitgeb's paradox is indeed an irreducible paradox of interaction whereas Halbach's paradox turns out to be a reducible paradox in this very sense.

The paper is structured as follows. We first briefly present the paradoxes of interaction and some further relevant facts. Next we introduce Smoryński's Diagonal Modal Logic and restate the paradoxes in this very setting. We then present the key conceptual innovation of this paper, namely the notion of reducibility we shall employ and immediately give an example of a reducible paradox. In a further section we introduce Kripke models for Diagonal Modal Logic and thereby provide a novel model existence theorem for a particular modal frame. In the last technical section we employ these semantic tools to show that Horsten and Leitgeb's paradox is irreducible. We end the paper with a brief philosophical evaluation of the results.

2 Paradoxes of Interaction

As we have already pointed out, if modalities are treated as predicates paradox may arise. In this section we shall present some of these paradoxes focussing on paradoxes of multiple modalities. We will start, however, by recalling Montague's [10] famous inconsistency result for one modal predicate which establishes that the assumption of two central modal principles leads to inconsistency. Let $\mathcal{L}_N$ be the language of arithmetic extended by a one-place predicate N. Montague showed that any $\mathcal{L}_N$-theory Σ extending Robinson arithmetic and which proves all instances of the axiom scheme

$$(T_N) \qquad N\ulcorner\phi\urcorner \to \phi$$

and which is closed under the rule

$$(Nec_N) \qquad \frac{\phi}{N\ulcorner\phi\urcorner}$$

is inconsistent. To produce his inconsistency result Montague used the fact that the so-called diagonal lemma is provable in any extension of Robinson arithmetic. Consequently, there is a sentence δ of $\mathcal{L}_N$ such that Σ proves

$$(\dagger) \qquad \neg N\ulcorner\delta\urcorner \leftrightarrow \delta$$

It is not hard to see that (†), (T_N) and (Nec_N) are jointly inconsistent if classical logic is assumed. Later in this paper we give a proof of Montague's theorem in a slightly different setting. At this point it suffices to observe the importance of the diagonal lemma in the derivation of the inconsistency result, that is the possibility of diagonalizing the modal predicate. As a matter of fact virtually all known modal paradoxes depend crucially on the application of the diagonal lemma and thus on diagonalizing the modal predicate.[1]

We now turn to the multimodal paradoxes, which have not seen that much attention so far. Notable exceptions are the work of Niebergall [12], Halbach [5, 6], and Horsten and Leitgeb [8]. In this paper we shall focus on the paradoxes presented by Halbach [6] and Horsten and Leitgeb [8] because in the two mentioned cases the paradox seems to arise from the interaction of the modal predicates. We wish to investigate whether this intuition can be substantiated.

In their paper "No Future" Horsten and Leitgeb [8] show that the principles of temporal modal logic lead to inconsistency if the two temporal modalities are treated as predicates. To see this consider a language $\mathcal{L}_{HG}$ which is an expansion of the language of arithmetic by two

[1] "Paradoxes" like Fitch's paradox form a notable exception to this claim. For an overview of the unimodal paradoxes obtained by appeal to diagonalization we refer the reader to Egré [4].

one-place predicates H, G. The intended reading of the predicate H is 'it always has been the case' and the intended reading of G is 'it is always going to be the case'.

Theorem 1 (Horsten and Leitgeb). *Let Σ be a $\mathcal{L}_{HG}$-theory extending Robinson arithmetic which proves all instances of the schemata*

(K_G)	$G(\ulcorner\phi\urcorner) \wedge G(\ulcorner\phi \rightarrow \psi\urcorner) \rightarrow G(\ulcorner\psi\urcorner)$
(K_H)	$H(\ulcorner\phi\urcorner) \wedge H(\ulcorner\phi \rightarrow \psi\urcorner) \rightarrow H(\ulcorner\psi\urcorner)$
(D_H)	$H(\ulcorner\phi\urcorner) \rightarrow \neg H(\ulcorner\neg\phi\urcorner)$
(In_{HG})	$\phi \rightarrow H(\ulcorner\neg G\ulcorner\neg\phi\urcorner\urcorner)$

and is closed under the following rules of necessitation

(Nec_G) $\dfrac{\phi}{G\ulcorner\phi\urcorner}$ $\qquad$ (Nec_H) $\dfrac{\phi}{H\ulcorner\phi\urcorner}.$

Then Σ is inconsistent.

Proof. By diagonalization there is a δ such that $\Sigma \vdash \delta \leftrightarrow H(\ulcorner\neg G\ulcorner\neg\neg\delta\urcorner\urcorner)$.

1.	$\neg\delta \leftrightarrow \neg H\ulcorner\neg G\ulcorner\neg\neg\delta\urcorner\urcorner$	*diag*
2.	$\neg\delta \rightarrow H\ulcorner\neg G\ulcorner\neg\neg\delta\urcorner\urcorner$	(In_{HG})
3.	$\neg\neg\delta$	$1, 2$
4.	$G\ulcorner\neg\neg\delta\urcorner$	$3, (Nec_G)$
5.	$H\ulcorner G\ulcorner\neg\neg\delta\urcorner\urcorner$	$4, (Nec_H)$
6.	$\neg H\ulcorner\neg G\ulcorner\neg\neg\delta\urcorner\urcorner$	$5, (D_H)$
7.	$H\ulcorner\neg G\ulcorner\neg\neg\delta\urcorner\urcorner$	$3,$ *diag*

□

Horsten and Leitgeb's inconsistency result seems to arise from the specific interaction of the forward and the backward looking temporal modality which is characterized by the multimodal principle (In_{HG}). (In_{HG}) asserts that whenever ϕ is the case, then it has always been the case that at some point in the future ϕ was going to be the case.

Again the derivation of the inconsistency hinges crucially on the application of the diagonal lemma. But in the present case the formula that we diagonalized contained both modal predicates. This sets the derivation of the inconsistency result apart from the unimodal paradoxes where formulas containing only one modal predicate are diagonalized. This observation will prove crucial later as, in a nutshell, we shall call a multimodal paradox a paradox of interaction proper if the derivation of the paradox requires diagonalizing a formula containing both modal notions.

Halbach [6] also proves his inconsistency result by diagonalizing a formula containing two modal predicates. Halbach is concerned with a predicate expressing an alethic modality and a predicate expressing the notion of knowledge. He shows that if the principle of knowability, i.e. Fitch's principle, is assumed in the predicate setting, then very basic principles for the alethic modality and knowledge lead to inconsistency. Thus in contrast to the operator setting where Fitch's principle leads to a collapse of knowledge and truth, we even obtain an inconsistency proper in the predicate setting. Let $\mathcal{L}_{KN}$ be an extension of the language of arithmetic by two one place predicates K and N. The predicate K is read as 'is known' while the predicate N is read as 'is necessary'.

Theorem 2 (Halbach)**.** *Let Σ be a $\mathcal{L}_{KN}$-theory extending Robinson arithmetic which proves all instances of the schemata*

$$(T_K) \qquad K\ulcorner\phi\urcorner \to \phi$$

$$(In_{NK}) \qquad \phi \to \neg N\ulcorner\neg K\ulcorner\phi\urcorner\urcorner$$

and which is closed under the following rule of necessitation

$$(Nec_N) \qquad \frac{\phi}{N\ulcorner\phi\urcorner}.$$

Then Σ is inconsistent.

At this point we shall not give a detailed proof of Halbach's result. Again, later in this paper, we present a proof of Halbach's result in a slightly different setting. For now it suffices to notice that Halbach uses the following instance of the diagonal lemma in deriving his result.

$$(\ddagger) \qquad \delta \leftrightarrow N\ulcorner\neg K\ulcorner\delta\urcorner\urcorner$$

That is, as we have indicated, Halbach diagonalizes a formula containing both modal predicates to establish his inconsistency result.

At least prima facie, Halbach's inconsistency result also seems to arise from the interaction of the two modal predicates and, in particular, from the principle (In_{NK}). Since Halbach diagonalizes a formula containing two modal predicates, distinguishing multimodal paradoxes depending on which formulas need to be diagonalized in deriving the inconsistency result looks like a promising way to go. As we have already mentioned, the idea would be that a multimodal paradox is a paradox of interaction proper if we need to diagonalize a formula containing multiple modal predicates to derive the inconsistency result. If the inconsistency can be derived by diagonalizing a formula containing only one modal predicate, we say that the paradox is reducible to the unimodal paradoxes.

Yet, it proves difficult to render this idea precise within the predicate setting. In order to show that a paradox is a paradox of interaction proper, i.e. irreducible to the unimodal paradoxes, one needs to show that the theory under consideration would be consistent if we were to diagonalize formulas containing one modal predicate only. But since the modal theories we consider extend Robinson arithmetic we can prove the diagonal lemma and thus diagonalize all formulas. It is unclear how the diagonal lemma can be restricted to allow for diagonalization of formulas containing just one modal predicate only. Moreover, such a restriction of the diagonal lemma would seem highly artificial.

There is a further reason why it appears difficult to spell out the aforementioned distinction within the predicate setting, and this reason actually reveals a sloppiness in our terminology. The problem is that within the predicate approach the formulas that were diagonalized in order to derive Theorems 1 and 2 do not really contain both modal predicates. Rather they *use* one predicate but only *mention* the second one. This might cause several complications in spelling out the distinction we are after in a precise way even though, intuitively, the guiding idea should be pretty clear.

We will therefore take a different route and, in fact, move back to the operator setting to render our idea precise. We will consider modal logics in which so-called fixed-points of the modal formulas are available. In the predicate setting a fixed-point is a sentence we obtain by diagonalizing a particular formula. In the operator approach fixed-points for all modal formulas are usually not available.[2] Nonetheless we can introduce these fixed-points to the language and logic which will allow us to spell out the modal paradoxes within the operator setting.[3] But in contrast to the predicate setting we are in full control which fixed-points we introduce to the framework and this will allow us to spell out our idea in a precise way. In the next section we introduce modal operator logic with fixed-points and then present the paradoxes of interaction in the modified setting.

3 Diagonal Modal Logic

Smoryński [13] introduced modal operator logics with fixed-points in studying the modal properties of provability. He introduced new operators to the language which applied to a formula provide the fixed-point of the formula. He labeled this extension of modal operator logic "Diagonalization Operator Logic" (DOL). In what is to come we shall consider a variant

[2]The modal operator logic *GL* ("Gödel-Löb") forms a notable exception to this rule. Sambin and De Jongh showed independently that *GL* has the fixed-point property, i.e. that we can find fixed-points for all modal formulas.

[3]Egré [4] already uses modal operator logics with fixed-points to provide a systematization and an overview of the (uni)modal paradoxes.

of (DOL) which dispenses with fixed-point operators in place of fixed-point constants. We shall call this variant of (DOL), which was also introduced by Smoryński [14], 'Diagonal Modal Logic" (DML).

As we just noted, Diagonal Modal Logic (DML) extends simple modal logic in that any system of DML will have the fixed-point property, independently of the modal principles assumed. That is, for every formula $\phi(p)$ in which p occurs only within the scope of a modal operator there will be a fixed-point. To this end new propositional constants will be introduced into the language, namely for every modal formula $\phi(p)$ with p boxed in ϕ, we introduce the fixed-point constant $\delta_{\phi(p)}$. A propositional atom p is boxed in ϕ, if and only if, all occurrences of p in ϕ are in the scope of at least one modal operator.[4] This will be achieved by considering a sequence of languages. At each step in the sequence we add new fixed-point constants for the formulas of the language in which a propositional variable occurs in the scope of a modal operator only. Eventually we take the language of DML to be the union of all members of the sequence. This way of construing the language guarantees that we add fixed-point constants for all formulas of the language in which a propositional constant occurs boxed. Otherwise we would not be guaranteed that for a formula, say, $\Box(p \wedge \delta_{\psi(q)})$ we can find an appropriate fixed-point because the fixed-point constant $\delta_{\Box(p\wedge\delta_{\psi(q)})}$ is not added at the first step in the sequence. The reason for this is that the formula $\Box(p \wedge \delta_{\psi(q)})$ is not a formula of the basic language of modal operator logic we shall now define.

Definition 3 (Modal Operator Language)**.** *The (uni)modal operator language $\mathcal{L}_\Box$ consists of a denumerable set of propositional atoms (propositional variables) $At_\Box$, a propositional constant $\bot$, the boolean operators $\neg$ and $\wedge$, and an one-place modal operator $\Box$. The notion of a well formed formula is defined by*

$$\phi ::= p \mid \bot \mid \neg\phi \mid \phi \wedge \phi \mid \Box\phi$$

with $p \in At_\Box$.

Before we give the definition of the language of DML we give a precise definition of when a propositional variable is boxed in a modal formula.

Definition 4. *Let $\phi(p)$ be a formula. If p occurs in $\phi(p)$, we say that p is boxed in $\Box\phi(p)$. If p does not occur in $\phi(p)$, we say that p is boxed in $\phi(p)$. Moreover, if p is boxed in ϕ and ψ, then p is boxed in $\neg\phi$, $\neg\psi$, $\Box\phi$, $\Box\psi$ and $\phi \wedge \psi$. Otherwise, p is not boxed in ϕ. We write $B(p, \phi)$ to convey the fact that p occurs and is boxed in ϕ.*

[4] Without the restriction to boxed propositional variables the approach would be trivialized. Diagonalizing the formula $\neg p$ would lead directly into contradiction. Smoryński [13] (pp. 72/73) points out that the restriction to boxed formulas is also well motivated.

The modal fixed-point language $\mathcal{L}_\Box^F$ is obtained from a modal operator language $\mathcal{L}_\Box$ by consecutively adding new propositional atoms to the language and at each stage closing off under the usual formation rules.

Definition 5 (Modal Fixed-Point Language)**.** *We set $L_0 = \mathcal{L}_\Box$ and thus $At_0 = At_\Box$ and define*

$$At_{n+1} = At_n \cup \{\delta_{\phi(p)} : \phi(p) \in L_n \,\&\, p \text{ is boxed in } \phi\}.$$

The language L_{n+1} is obtained by closing off At_{n+1} by the ordinary formation rules of $\mathcal{L}_\Box$. That is, the syntax of L_{n+1} is as follows

$$\phi := c \mid \bot \mid \neg\phi \mid \Box\phi \mid \phi \wedge \phi$$

with $c \in At_{n+1}$. Eventually, we set

$$At_\Box^F = \bigcup_{n\in\omega} At_n \ ;$$
$$\mathcal{L}_\Box^F = \bigcup_{n\in\omega} L_n \ .$$

Having defined the language we can move on to the more important task of giving a precise definition of 'Diagonal Modal Logic' which will be the extension of a modal operator logic in the language $\mathcal{L}_\Box^F$ by suitable fixed-point axioms. We assume modal operator logics to be presented using axiom schemata. As a consequence we assume that the rule of uniform substitution is not required in axiomatizing modal operator logics. This is important as Diagonal Modal Logic will not be closed under the rule of uniform substitution.[5]

Definition 6 (Diagonal Modal Logic)**.** *Let $\mathcal{S}$ be a modal operator logic in the language $\mathcal{L}_\Box^F$. The Diagonal Modal Logic $\mathcal{S}^F$ is obtained from $\mathcal{S}$ by adding the fixed-point axioms*

$$\phi(\delta_{\phi(p)}) \leftrightarrow \delta_{\phi(p)}$$

for all formulas $\phi(p)$ with $B(p,\phi)$ to the logic $\mathcal{S}$. We call $\mathcal{S}^F$ the diagonal extension of $\mathcal{S}$.

Diagonal Modal Logic allows us to restate most inconsistency results of the predicate setting, such as Montague's theorem, which are derived using the diagonal lemma in a straightforward way.

[5]This suggests that an operator formulation of DML along the lines of DOL or Alberucci and Facchini's [1] modal $\mu^{\sim}$-calculus may be preferable. Still, in many respects the present formulation in terms of fixed-point constants proves sufficient and is somewhat easier to handle. It may also be preferable not to call DML a logic but rather a theory, however with the mentioned caveat we stick to the original terminology.

Theorem 7 (Montague). *Let $\mathcal{S}$ be a modal operator logic in $\mathcal{L}_\Box$. If $\mathcal{S}$ is axiomatized by all instances of*

$$(T_\Box) \qquad \Box\phi \to \phi$$

and the rule of necessitation. Then $\mathcal{S}^F$ is inconsistent in $\mathcal{L}^F_\Box$.

Proof. The proof follows the pattern of Montague's original proof.

1. $\neg\Box\delta_{\neg\Box p} \leftrightarrow \delta_{\neg\Box p}$ — *DML*
2. $\Box\delta_{\neg\Box p} \to \delta_{\neg\Box p}$ — (T)
3. $\neg\Box\delta_{\neg\Box p}$ — 1, 2
4. $\delta_{\neg\Box p}$ — 1, 3
5. $\Box\delta_{\neg\Box p}$ — 4, $(Nec_\Box)$

$\Box$

After this brief presentation of DML in the unimodal setting we return to a framework that allows for multiple modalities. As the paradoxes under consideration will involve only two modalities, we confine ourselves to multimodal logics with two modal operators $\Box$ and $\blacksquare$. Let $\mathcal{L}_\Box$ and $\mathcal{L}_\blacksquare$ be the two corresponding unimodal operator languages and $\mathcal{L}_{\Box\blacksquare}$ the multimodal operator language whose syntax is given by

$$\phi ::= p \mid \bot \mid \neg\phi \mid \phi \wedge \phi \mid \Box\phi \mid \blacksquare\phi$$

where $p \in At_{\Box\blacksquare}$. For sake of convenience we assume the three modal languages to agree on their propositional atoms. Now, there will be two ways how we may obtain multimodal fixed-point languages extending $\mathcal{L}_{\Box\blacksquare}$ which will differ on the fixed-point constants available.

One option is to define the multimodal fixed-point language $\mathcal{L}^F_{\Box\blacksquare}$, which we may obtain from a multimodal language $\mathcal{L}_{\Box\blacksquare}$ along the lines of Definition 5. To this end we need to define what it means for p to be boxed in ϕ in the multimodal setting.

Definition 8. *Let $\phi(p)$ be a formula. If p occurs in $\phi(p)$, we say that p is boxed in $\Box\phi(p)$ and we also say that p is boxed in $\blacksquare\phi(p)$. If p does not occur in $\phi(p)$, we say that p is boxed in $\phi(p)$. Moreover, if p is boxed in ϕ and ψ, then p is boxed in $\neg\phi$, $\neg\psi$, $\Box\phi$, $\Box\psi$, $\blacksquare\phi$, $\blacksquare\psi$ and $\phi \wedge \psi$. Otherwise, p is not boxed in ϕ. We write $B(p, \phi)$ to convey the fact that p occurs and is boxed in ϕ.*

Given this definition, it is easily seen how to adopt Definition 5 in order to obtain the multimodal fixed-point language $\mathcal{L}^F_{\Box\blacksquare}$ from $\mathcal{L}_{\Box\blacksquare}$. This language is to be distinguished from $\mathcal{L}^F_\Box \cup \mathcal{L}^F_\blacksquare$, the join of the two modal fixed-point languages $\mathcal{L}^F_\Box$ and $\mathcal{L}^F_\blacksquare$ which we take to be the modal fixed-point languages in the sense of Definition 5 obtained from $\mathcal{L}_\Box$ and $\mathcal{L}_\blacksquare$ respectively. The syntax of $\mathcal{L}^F_\Box \cup \mathcal{L}^F_\blacksquare$ is given by:

$$\phi := c \mid \bot \mid \neg\phi \mid \Box\phi \mid \blacksquare\phi \mid \phi \wedge \phi$$

with $c \in At^F_\Box \cup At^F_\blacksquare$.

The crucial difference between the two multimodal fixed-point languages $\mathcal{L}^F_{\Box\blacksquare}$ and $\mathcal{L}^F_\Box \cup \mathcal{L}^F_\blacksquare$ is that the former contains fixed-point constants of formulas in which both modal operators occur whereas the latter does not, e.g. $\delta_{\Box\blacksquare p}$ is a propositional constant of the former but not of the latter language. We therefore distinguish between two diagonal extensions of a multimodal logic $\mathcal{S}$: the diagonal extension in the language $\mathcal{L}^F_{\Box\blacksquare}$ and the diagonal extension in the language $\mathcal{L}^F_\Box \cup \mathcal{L}^F_\blacksquare$ where the notion of a diagonal extension (or a DML) is modified to the effect that we only introduce fixed-point axioms for all fixed-point constants available in the language under consideration (not for all formulas in which p occurs boxed in ϕ).

The distinction between these two different diagonal extensions of a multimodal logic will be of central importance in distinguishing between the reducible and irreducible multimodal paradoxes. But before we provide the precise criterion of reducibility, we state the modal paradoxes in the setting of DML.

We now reconsider the paradoxes of interaction in the framework of DML. We shall present three paradoxes of interaction. Two of these paradoxes, namely the one by Horsten and Leitgeb, and the one by Halbach, we have already presented in Section 2. This time we start by presenting Halbach's result and provide a proof of his result in the novel setting.

Theorem 9 (Halbach)**.** *Let $\mathcal{S}$ be a multimodal logic in $\mathcal{L}_{\Box\blacksquare}$. If $\mathcal{S}$ is axiomatized by all instances of*

($\blacksquare T$)	$\blacksquare\phi \rightarrow \phi$
($\Diamond\blacksquare In$)	$\phi \rightarrow \Diamond\blacksquare\phi$

and the rule

($\Box Nec$) $$\frac{\phi}{\Box\phi}$$

then $\mathcal{S}^F$ is inconsistent in $\mathcal{L}^F_{\Box\blacksquare}$.

Proof. The proof is due to Halbach [6].[6]

1. $\delta \leftrightarrow \neg\Diamond\blacksquare\delta$ — *DML*
2. $\delta \rightarrow \Diamond\blacksquare\delta$ — ($\Diamond\blacksquare In$)
3. $\neg\delta$ — 1,2

[6]For ease of presentation we simply write δ instead of the correct $\delta_{\neg\Diamond\blacksquare p}$ throughout the proof.

4. $\blacksquare\delta \to \delta$ $\hfill (\blacksquare T)$
5. $\neg\blacksquare\delta$ $\hfill 3,4$
6. $\Box\neg\blacksquare\delta$ $\hfill 5, (\Box Nec)$
7. $\neg\Diamond\blacksquare\delta$ $\hfill 6, Def.$
8. δ $\hfill 1,7$

$\Box$

Next we state Horsten and Leitgeb's result in the new setting. We won't rehearse the proof we have given in the predicate setting but provide an alternative proof.

Theorem 10 (Horsten and Leitgeb)**.** *Let $\mathcal{S}$ be a multimodal logic in $\mathcal{L}_{\Box\blacksquare}$. If $\mathcal{S}$ is axiomatized by*

$(\Box K)$	$\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$
$(\blacksquare K)$	$\blacksquare(\phi \to \psi) \to (\blacksquare\phi \to \blacksquare\psi)$
$(\Box D)$	$\Box\phi \to \Diamond\phi$
$(\Box\blacklozenge In)$	$\phi \to \Box\blacklozenge\phi$

and the rules $(\Box Nec)$ *and* $(\blacksquare Nec)$*, then $\mathcal{S}^F$ is inconsistent in $\mathcal{L}^F_{\Box\blacksquare}$.*[7]

As mentioned, instead of proving Horsten and Leitgeb's result along the lines of Theorem 1, we use the next theorem to this effect. The theorem adopts Montague's theorem to the multimodal setting.

Theorem 11 (Multimodal Montague)**.** *Let $\mathcal{S}$ be a multimodal logic in $\mathcal{L}_{\Box\blacksquare}$. If $\mathcal{S}$ is axiomatized by*

$(\Box\blacksquare T)$	$\Box\blacksquare\phi \to \phi$

and the rules $(\Box Nec)$ *and* $(\blacksquare Nec)$*, then $\mathcal{S}^F$ is inconsistent in $\mathcal{L}^F_{\Box\blacksquare}$.*

Proof. The proof is a small modification of Montague's original argument in [10].[8]

1. $\delta \leftrightarrow \neg\Box\blacksquare\delta$ $\hfill DML$
2. $\Box\blacksquare\delta \to \delta$ $\hfill (\Box\blacksquare T)$
3. δ $\hfill 1,2$
4. $\Box\blacksquare\delta$ $\hfill 3, (\blacksquare Nec), (\Box Nec)$
5. $\neg\Box\blacksquare\delta$ $\hfill 1,3$

$\Box$

[7] $(\blacksquare Nec)$ is just the rule of necessitation for the $\blacksquare$-operator.

[8] Again we omit the subscript to the fixed point constant $\delta_{\neg\Box\blacksquare p}$ throughout the proof.

Interestingly, upon closer inspection Theorem 10 appears to be a corollary of the preceding theorem, for ($\Box\blacksquare T$) is derivable in the multimodal logic $\mathcal{S}$ assumed in Theorem 11.

Lemma 12. *Let $\mathcal{S}$ be a multimodal logic in $\mathcal{L}_{\Box\blacksquare}$ as specified in Theorem 10. Then $\mathcal{S} \vdash (\Box\blacksquare T)$.*

Proof. We derive ($\Box\blacksquare T$) in $\mathcal{S}$

1. $\neg\phi \to \Box\blacklozenge\neg\phi$ $\quad (\Box\blacklozenge In)$
2. $\neg\phi \to \Box\neg\blacksquare\neg\neg\phi$ $\quad 1, Def$
3. $\neg\phi \to \Box\neg\blacksquare\phi$ $\quad (\Box/\blacksquare Nec), (\Box/\blacksquare K)$
4. $\neg\phi \to \neg\Box\blacksquare\phi$ $\quad 3, (\Box D)$
5. $\Box\blacksquare\phi \to \phi$ $\quad 4$

$\Box$

We can now provide an easy proof of Theorem 10.

Proof of Theorem 10. By Lemma 12 we know that $\mathcal{S} \vdash (\Box\blacksquare T)$ and thus, by Theorem 11, $\mathcal{S}^F$ is inconsistent in $\mathcal{L}^F_{\Box\blacksquare}$. $\Box$

After this outline of Diagonal Modal Logic and our presentation of the multimodal paradoxes in the setting of DML, we return to our initial question of whether these are genuine paradoxes of interaction or whether they are reducible in some pertinent sense to the unimodal paradoxes.

4 Reducibility

We now return to our initial question: Are these genuinely new paradoxes, i.e. genuine paradoxes of interaction, or can they be "reduced" in some way to the paradoxes we discussed in the unimodal setting?

The idea behind this question is that we can draw a distinction along the following lines. On the one hand we have paradoxes where due to a new modal axiom, possibly an axiom of interaction, we have strengthened one of the unimodal sublogics of the multimodal logic to the effect that its diagonal extension will be inconsistent. On the other hand, there might be proper paradoxes of interaction for which the diagonal extensions of the unimodal logics at stake are consistent whereas the diagonal extension of the multimodal logic is inconsistent. We shall call the former paradoxes reducible and the latter ones irreducible.

Obviously, one has to be careful here because we are dealing with inconsistent multimodal logics. Let $\mathcal{S}^F$ be such an inconsistent multimodal logic. Then by the inconsistency of $\mathcal{S}^F$ and the *ex falso quodlibet* we know that all unimodal logics, also the inconsistent ones, are sublogics of $\mathcal{S}^F$. Thus the question arises of how to make this intuitive distinction precise. Exactly at this point the distinction between the different fixed-point languages and logics comes in handy because it seems that the reducible paradoxes should be derivable by diagonalizing a formula containing only one modality whereas this should not be possible if we are dealing with irreducible paradoxes. Accordingly, we claim to be in the first case, that is in the reducible case, if we can derive the inconsistency result using a fixed-point axiom of a DML, say $\mathcal{S}^F$, formulated in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$ and in the latter case, the irreducible one, if we need to retort to a fixed-point axiom available only when $\mathcal{S}^F$ is formulated in $\mathcal{L}^F_{\Box\blacksquare}$. This way of drawing the distinction between reducible and irreducible paradoxes also seems to fit the observation we made earlier in this paper, namely that the derivations of the paradoxes we have presented all seem to require the diagonalization of a formula involving both modal notions at stake.

Definition 13 (Reducibility and Irreducibility of Inconsistencies)**.** *Let $\mathcal{S}^F$ be a DML inconsistent in $\mathcal{L}^F_{\Box\blacksquare}$. We say that the inconsistency of $\mathcal{S}^F$ or the paradox is* reducible *to the unimodal case iff $\mathcal{S}^F$ is inconsistent in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$. We say the paradox is* irreducible *iff $\mathcal{S}^F$ is consistent in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$.*

According to our criterion a multimodal paradox will be irreducible if the diagonal extension of the multimodal logic is consistent as long as no fixed-points for multimodal formulas are available. A paradox will be called reducible if the underlying multimodal logic is inconsistent even in the absence of fixed-points of the multimodal formulas.[9]

The remainder of the paper is devoted to showing that this criterion actually manages to draw a distinction between different multimodal paradoxes. That is, we will show that Halbach's paradox turns out to be a reducible paradox whereas Horsten and Leitgeb's paradox proves to be an irreducible paradox. Establishing the former claim is relatively straightforward and we shall provide the proof now. However showing that Horsten and Leitgeb's paradox is irreducible will require some more work.

As promised, we now show that Halbach's paradox turns out to be a reducible paradox. To this end we show that the modal principles Halbach used in deriving his inconsistency result are already jointly inconsistent in the language $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$.

[9] A referee pointed to the fact that the reducibility of a paradox could also be spelled out in terms of conservativity: a paradox is reducible, if and only if, $\mathcal{S}^F$ in $\mathcal{L}^F_{\Box\blacksquare}$ is $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$-conservative over $\mathcal{S}^F$ in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$. While this is true, spelling out the criterion in this way does not provide further insight since we always start with $\mathcal{S}^F$ inconsistent in $\mathcal{L}^F_{\Box\blacksquare}$. As a consequence whenever $\mathcal{S}^F$ is inconsistent in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$, $\mathcal{S}^F$ in $\mathcal{L}^F_{\Box\blacksquare}$ will be a $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$-conservative extension. Whenever $\mathcal{S}^F$ in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$ is consistent, then $\mathcal{S}^F$ in $\mathcal{L}^F_{\Box\blacksquare}$ will not be a conservative extension. For a discussion of the notion of conservative extension in the setting of multimodal logics see, for example, Williamson [15].

Theorem 14. *Let $\mathcal{S}$ be a multimodal logic in $\mathcal{L}_{\Box\blacksquare}$. If $\mathcal{S}$ is axiomatized by* ($\blacksquare T$), ($\Diamond\blacksquare In$) *and the rule* ($\Box Nec$), *then $\mathcal{S}^F$ is inconsistent in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$.*

Proof. The proof uses a fixed-point of the language $\mathcal{L}^F_{\blacksquare}$

1. $\delta_{\neg\blacksquare p} \leftrightarrow \neg\blacksquare\delta_{\neg\blacksquare p}$ — DML
2. $\blacksquare\delta_{\neg\blacksquare p} \rightarrow \delta_{\neg\blacksquare p}$ — $(\blacksquare T)$
3. $\delta_{\neg\blacksquare p}$ — $1, 2$
4. $\neg\blacksquare\delta_{\neg\blacksquare p}$ — $1, 3$
5. $\Box\neg\blacksquare\delta_{\neg\blacksquare p}$ — $4, (\Box Nec)$
6. $\neg\Diamond\blacksquare\delta_{\neg\blacksquare p}$ — $5, Def.$
7. $\delta_{\neg\blacksquare p} \rightarrow \Diamond\blacksquare\delta_{\neg\blacksquare p}$ — $(\Diamond\blacksquare In)$
8. $\Diamond\blacksquare\delta_{\neg\blacksquare p}$ — 3

□

Corollary 15. *Halbach's paradox is reducible in the sense of Definition 13.*

This establishes that there are reducible multimodal paradoxes. Next we turn to showing that there are also irreducible paradoxes. As pointed out, we show Horsten and Leitgeb's inconsistency result to be irreducible in the sense of Definition 13. However, before we can show this, we need to introduce Kripke models for Diagonal Modal Logic, which we do in the next section.

5 Models for Diagonal Modal Logic

In the previous sections we investigated DMLs from a proof-theoretic perspective. In this section we turn to more semantic questions and provide Kripke models for certain DMLs. As Kripke models and normal modal logics are at the heart of standard modal operator logic this seems to be a good starting point for an investigation into semantic aspects of DML. Moreover Kripke models for DML will prove sufficient for our purpose as we shall be dealing with diagonal extensions of normal modal logics. The question then is which Kripke models for modal operator logic can be extended to models for DML. More precisely, the question is which models of a given modal logic $\mathcal{S}$ can be extended into a model for $\mathcal{S}^F$. Ideally, we would like a (constructive) procedure that allows us to extend an appropriate Kripke model of some normal modal logic $\mathcal{S}$ to a model of the diagonal extension of $\mathcal{S}$.

Fortunately, it turns out that such a procedure is available for a wide class of models. Smoryński [13] showed that all Kripke-models based on a converse well-founded frame can be turned into models for DML.[10] This result implies that if $\mathcal{S}$ is a modal logic valid on a converse well-founded frame $\mathcal{F}$, M a model based on this frame, then M can be extended to a model M' of $\mathcal{S}^F$. Before we give the details of Smoryński's construction we define the notion of 'converse well-foundedness'.

Definition 16 (Converse well-founded)**.** *A relation R is converse well-founded on a set W iff for all nonempty $X \subseteq W$ there exists a $w \in X$ s.t. for all $v \in X, \neg(wRv)$.*

The reason why we can provide interpretations for the fixed-point constants on converse well-founded frames is that each R-sequence will eventually hit a dead end, i.e. a world that sees no world (has no accessible worlds). In these worlds the interpretation of the fixed-point constants does not matter for the valuation of the fixed-point formula and thus we can simply give the fixed-point constant the truth value of the fixed-point formula. Then we can move on downwards through the model repeating this procedure, as we have already defined the interpretation of the fixed-point constants in the worlds relevant for the evaluation of the fixed-point formula.

To make this sketch precise we need to introduce the auxiliary notion of the "rank" of a world. Once this notion is at hand we can induct over the rank of a world.

Definition 17 (Rank)**.** *Let R be a converse well-founded dyadic relation on a set W. Then we can associate with each world w an ordinal α which is said to be the rank of w.*[11] *The rank $\rho(w)$ of a world w is defined inductively as follows:*

$$\rho(w) := \begin{cases} 0 & ,\textit{if} \quad \neg\exists v(wRv) \\ \alpha + 1 & ,\textit{if} \quad \exists v \in W(wRv \,\&\, \rho(v) = \alpha \,\&\, \forall u(wRu \Rightarrow \rho(u) \leq \alpha)) \end{cases}$$

Note that since R is converse well-founded on W we know that for every $w \in W$ a dead end is reached in finitely many steps and furthermore that no infinite R-path starts from w.[12]

In the proof of the following theorem we shall sketch how to construct models for $\mathcal{L}_\Box^F$ out of models of $\mathcal{L}_\Box$ based on converse well-founded frames.

Theorem 18 (Smoryński)**.** *Let $\mathcal{F} = \langle W, R\rangle$ be a converse well-founded frame, $(\mathcal{F}, V)$ a $\mathcal{L}_\Box$-model with a valuation function V. Then there exists a $\mathcal{L}_\Box^F$-model $(\mathcal{F}, V^F)$ with $V \subset V^F$ such that for all*

[10] The reader acquainted with provability logic and the fixed-point theorem for *GL* by De Jongh and Sambin might note that the semantic result is actually stronger as *GL* is not valid on all converse well-founded frames.

[11] For the notion of 'rank' cf. Boolos [3], pp. 94-95.

[12] For assume to the contrary that there is a world w from which an infinite R-path starts. Define a subset X of W as the set of those worlds that can be reached in finitely many steps from w and which lie on the infinite R-path. By converse well-foundedness there needs to be an R-maximal element in X which contradicts the initial assumption. This also implies that a dead end is always reached in finitely many steps.

fixed-point constants $\delta_{\phi(p)}$ of $\mathcal{L}_{\Box}^{F}$

$$(\mathcal{F}, V^F) \models \phi(\delta_{\phi(p)}) \leftrightarrow \delta_{\phi(p)}.$$

Proof. By an induction over the rank of a world we can extend the valuation V to a suitable valuation V^F. That is, we first show how to give the interpretation of a fixed-point constant $\delta_{\phi(p)}$ at all worlds of rank 0 and then, assuming the interpretation of $\delta_{\phi(p)}$ to be defined at worlds of rank α, show how to give the interpretation on worlds of rank $\alpha+1$. In fact, besides this prior induction process we need to induce over the construction process of $\mathcal{L}_{\Box}^{F}$ as well. We shall ignore this second induction, which is actually the main induction, but it should be kept in mind. The construction follows the outlines of Smoryński [13], pp. 109-112.

Let w be a world with $\rho(w) = 0$. Since $\delta_{\phi(p)}$ only occurs boxed in $\phi(\delta_{\phi(p)})$ we can rewrite $\phi(\delta_{\phi(p)})$ as a truth-functional decomposition $\psi(\Box\chi_1(\delta_{\phi(p)}), \ldots, \Box\chi_n(\delta_{\phi(p)}), \theta_1, \ldots, \theta_m)$ where $\delta_{\phi(p)}$ only occurs in the components $\Box\chi_i(\delta_{\phi(p)})$. Since there is no world accessible from w the components, $\Box\chi_i(\delta_{\phi(p)})$, will be true whatever interpretation we give to the fixed-point constants in w. Since the interpretation of all other components of the truth-functional decomposition is already defined on $(\mathcal{F}, V)$ (note that at this point the second induction we mentioned comes in) we can define:

$$V_0(\delta_{\phi(p)}) := \{w \in W : (\mathcal{F}, V), w \models \phi(p) \,\&\, \rho(w) = 0\}.$$

For the induction step assume a world w of rank $\alpha + 1$ and for all worlds v with $\rho(v) < \rho(w)$ the interpretation of $\delta_{\phi(p)}$ is already defined. If we reconsider the truth-functional decomposition of $\phi(\delta_{\phi(p)})$ we can see that the truth value of the decomposition at a world w only hinges on the interpretation of $\delta_{\phi(p)}$ in the accessible worlds, and thus on worlds with rank smaller than w. This allows us to define:

$$V_{\alpha+1}(\delta_{\phi(\delta_{\phi(p)})}) := \{w \in W : (\mathcal{F}, V_\alpha), w \models \phi(\delta_{\phi(p)}) \,\&\, \rho(w) \leq \alpha + 1\}$$

Finally, we set

$$V^F(\delta_{\phi(p)}) := \bigcup_{\alpha \in \omega} V_\alpha(\delta_{\phi(p)}).$$

Clearly, $(\mathcal{F}, V^F) \models \phi(\delta_{\phi(p)}) \leftrightarrow \delta_{\phi(p)}$ and we are done. $\Box$

Definition 19. *A model M of $\mathcal{L}_{\Box}^{F}$ in which the fixed-point axioms are true is called a DML model. If M is a DML model and extends some model M' of $\mathcal{L}_{\Box}$, M is called a DML-extension of M'.*

Corollary 20. *Let $\mathcal{S}$ be some simple modal logic valid on some converse well-founded frame $\mathcal{F}$. Then $\mathcal{S}^F$ is true on any DML extension of an $\mathcal{L}_{\Box}$-model based on $\mathcal{F}$.*

Two properties of converse well-founded frames were crucial in the construction. Firstly, the fact that these frames have dead-ends was indispensable for the start of the induction, but secondly, the notion of the rank of a world was well defined, which allowed us to apply induction in the proof. These two properties of converse well-founded frames make it difficult to extend this construction to frames with other properties. For example, the notion of a rank is not well defined on any frame that contains an R-loop.

Still in some cases in which the underlying frame is not converse well-founded it is possible to construct DML-models. A case that will prove particularly interesting to us are models based on the frame (ω, S), i.e. the natural numbers ordered by the successor relation.

$$\begin{tikzcd} \overset{0}{\bullet} \arrow[r, "S"'] & \overset{1}{\bullet} \arrow[r] & \overset{2}{\bullet} \arrow[r] & \overset{3}{\bullet} \cdots\cdots \end{tikzcd}$$

In this frame each world only sees its immediate successor (and no other world). (ω, S) is a serial frame and consequently the modal operator logic KD will be valid on this frame. Although we cannot give a *constructive* method for how to extend any model based on (ω, S) to a DML-model, we can show that there exists a DML-extension for every model based on this frame.

Theorem 21. *Every model M based on (ω, S) can be extended to a DML model M' based on (ω, S).*

Before we give the proof of this theorem we outline the general idea behind the proof and prove some auxiliary lemmata. As in the case of the models based on converse well-founded frames we extend a valuation V of an $\mathcal{L}_\Box$-model based on (ω, S) to a valuation V' of the language $\mathcal{L}_\Box^F$. The resulting model (ω, S, V') will be a DML model, i.e. all the fixed-point axioms will be true at each $n \in \omega$.

Again the interpretation V' will be constructed by an induction on the construction process of $\mathcal{L}_\Box^F$ (cf. Definition 5). As induction hypothesis we assume that the language $\mathcal{L}_n$ is already interpreted. That is, we assume the existence of a valuation V_n that provides an interpretation for all propositional constants of At_n (again cf. Definition 5) and, moreover, assume that V_n verifies all fixed-point axioms for the fixed-point constants in At_n. We have to expand our valuation function V_n to a valuation V_{n+1} which provides an interpretation for the newly added propositional constants, i.e. the newly added fixed-point constants, and show that the fixed-point axioms will be satisfied at each world in ω. The valuation function V' will then be the union of the valuation functions V_n for $n \in \omega$.

In the frame (ω, S) a valuation for a propositional constant $\delta_{\psi(q)} \in At_{n+1}$ is a function $V_{\delta_{\psi(q)}} : \omega \to \{0, 1\}$. By the induction hypothesis we know that the formula $\psi(\delta_{\psi(q)})$ has a determinate truth value, if we provide the interpretation for $\delta_{\psi(q)}$ because ψ does not contain other fixed-point constants for which no interpretation has been provided so far. In other words the formula $\psi(q)$ can be viewed as a function $\psi^V_q : 2^\omega \to 2^\omega$ that takes the interpretation of $\delta_{\psi(q)}$ as an input and provides the interpretation of $\psi(\delta_{\psi(q)})$ as an output. What we are looking for is a fixed-point of this function, that is we are looking for an input of this function that has itself as an output. This fixed-point could then serve as the interpretation of $\delta_{\psi(q)}$ and the fixed-point axiom $\delta_{\psi(q)} \leftrightarrow \psi(\delta_{\psi(q)})$ will be true at each $n \in \omega$ under this interpretation. We are thus looking for a countably infinite binary sequence s such that $\psi^V_q(s) = s$.

If $s \in 2^\omega$ we write $V \cup s$ for the extension of V by the interpretation of $\delta_{\psi(q)}$, i.e. the countably infinite binary sequence s. We write $s[n, n+m]$ if we are only interested in the values of s in the interval between n and $n+m$, i.e. the values $(s)_n, (s)_{n+1}, \ldots, (s)_{n+m}$. Correspondingly, we write $V \cup s[n, n+m], k \models \phi$ if and only if ϕ is true at a world k in the frame (ω, S) for a valuation $V \cup t$, for all countably infinite binary sequences t, as long as $t_i = s_i$ for $n \leq i \leq n+m$. Thus if $V \cup s[n, n+m], k \models \psi(\delta_{\psi(q)})$, then the truth value of $\psi(\delta_{\psi(q)})$ at k only depends on the interpretation of $\delta_{\psi(q)}$ at the worlds in between n and $n+m$.

Now, since q is boxed in $\psi(q)$ the truth value of $\psi(\delta_{\psi(q)})$ at a world n in a model based on the frame (ω, S) will only depend on successor worlds of n, that is the interpretation of $\psi(\delta_{\psi(q)})$ at n is independent of the interpretation of $\delta_{\psi(q)}$ at n and any predecessor of n. Moreover since the modal depth of any modal formula, i.e the maximum number of modal embeddings in a formula, is finite, there will be a natural number m such that the interpretation of $\psi(\delta_{\psi(q)})$ at a world n will only depend on the interpretation of $\delta_{\psi(q)}$ in the m worlds subsequent to n. These observations lead to the following lemma:

Lemma 22. *For any $s \in 2^\omega$ and formulas ψ of modal depth m*

$$V \cup s, n \models \psi(\delta_{\psi(q)}) \Leftrightarrow V \cup s[n+1, n+m], n \models \psi(\delta_{\psi(q)}).$$

Consider an arbitrary sequence $s \in 2^\omega$. Then, by Lemma 22, for $\psi(q)$ with modal depth m the nth entry of the sequence $\psi^V_q(s)$, i.e. $(\psi^V_q(s))_n$ is determined by $(s)_{n+1}, \ldots, (s)_{n+m}$. So if we consider the sequence s' which is identical to s for all entries $(s)_i$ for $n+1 \leq i$ but where $(s')_i = (\psi^V_q(s))_i$ for $i < n+1$, then $(\psi^V_q(s'))_n = (s')_n$ and thus s' is a partial fixed-point of ψ^V_q, i.e. a fixed-point for the nth entry. We shall apply this argument repeatedly in proving the following generalization:

Lemma 23. *For all $n \in \omega$ and formulas ψ of modal depth m there is a sequence $s \in 2^\omega$ such that for all $i \leq n$, $(s)_i = (\psi^V_q(s))_i$.*

Proof. We take an arbitrary $n \in \omega$ and an arbitrary countably infinite binary sequence s_0. We define a new sequence s_1 as follows. Let $(s_1)_i = (s_0)_i$ for $n < i$ and set $(s_1)_i = (\psi^V_q(s_0))_i$ for $i \leq n$.

The sequence s_1 is a fixed-point for the entry n of the function ψ_q^V by the argument we have sketched above. Next we define a sequence s_2 which is a fixed-point of both, the entry n and the entry $n-1$. To this end set $(s_2)_i = (s_1)_i$ for $n-1 < i$ and $(s_2)_j = (\psi_q^V(s_1))_j$ for $j \leq n-1$. s_2 is the desired fixed-point for n and $n-1$. It is a fixed-point for n because the entries of s_2 in between $n+1$ and $n+m$ are the same as in s_1 and it is a fixed-point of $n-1$ by the argument we have appealed to before. Now clearly we can repeat this process and construct sequences $s_3, s_4, \ldots, s_n$ which will be partial fixed-points of the function ψ_q^V. The last sequence we can construct, i.e. s_n, will be a fixed-point for the entries from $n-(n-1)$ to n, that is the first n entries. We thus have $(s_n)_i = (\psi_q^V(s_n))_i$ for all $i \leq n$ and we have found a countably infinite binary sequence s such that for all $i \leq n$, $(s)_i = (\psi_q^V(s))_i$.

□

The lemma establishes that for all $n \in \omega$ there is a sequence that has the fixed-point property for the first n entries. But for any sequence s with the fixed-point property for the first n entries we also have the fixed-point property for all initial segments of s of length less than n. This allows us to define a binary branching tree $(T_{\psi_q^V}, <)$ as follows: Let $B := \bigcup_{i \leq \omega} 2^i$. If s is some sequence and $m \leq lh(s)$ then $s \upharpoonright m$ is the restriction of s to the first m entries. Now, define

$$T_{\psi_q^V} := \{s \in B : \text{there is a sequence } s' \in 2^\omega \text{ such that } s = s' \upharpoonright lh(s) \text{ and } \psi_q^V(s') \upharpoonright lh(s) = s\}$$

with

$$s < t :\Leftrightarrow lh(s) < lh(t) \,\&\, s = t \upharpoonright lh(s).$$

for $s, t \in T_{\psi_q^V}$. The tree is clearly binary branching and since we have initial segments of arbitrary length the tree is infinite. Therefore, by Weak König's lemma this tree has an infinite path and this path will be the fixed-point of the function ψ_q^V.

Lemma 24. *(Weak König's Lemma) Let $(T, <)$ be an infinite binary branching tree. Then T contains an infinite path.*

Finally, we have all the necessary prerequisites for proving Theorem 21.

Proof of Theorem 21. Let M be a model based on the frame (ω, S) and with an evaluation function V. Let $\delta_{\psi(q)}$ be a new constant introduced at stage $n+1$. The corresponding tree $T_{\psi_q^V}$ is infinite and binary branching. By Weak König's Lemma it contains an infinite path. By the definition of $T_{\psi_q^V}$ this path is a binary sequence s with the fixed-point property, i.e. $\psi_q^V(s) = s$. We can expand V to $V \cup s$. Since the fixed-point constants introduced in each step are independent of each other we can expand our valuation V by the relevant fixed-point sequences. As this works for arbitrary n we get our desired valuation V' as the union of all these valuations. $M' = (\omega, S, V')$ is a DML model based on (ω, S). □

Corollary 25. *Let $\mathcal{S}$ be the normal modal logic axiomatized by* (*K*), (*D*) *and the rule* (*Nec*) *and let M be an $\mathcal{L}_{\Box}$-model based on (ω, S). Then $\mathcal{S}^F$ is true in any DML-extension of M.*

The corollary of Theorem 21 guarantees that we can construct a *DML*-model in which the modal principles for the □-operator, which were assumed in deriving Horsten and Leitgeb's inconsistency result, will be true. This will play an important role in showing that Horsten and Leitgeb's multimodal paradox is irreducible in the sense of Definition 13.

6 An Irreducible Paradox of Interaction

In what is to come we show that the inconsistency result of Horsten and Leitgeb is genuine, i.e. irreducible to the unimodal paradoxes. This result is interesting as it confirms our initial claim that new and unexpected paradoxes may arise once we let modalities interact. That is, in order to avoid the paradoxes, it will not be sufficient to just keep track of the unimodal sublogics of the multimodal logic and to guarantee that their diagonal extensions are consistent.

According to our definition an inconsistency is irreducible if and only if the diagonal extension $\mathcal{S}^F$ of a multimodal logic $\mathcal{S}$ is consistent in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$. In order to show Horsten and Leitgeb's inconsistency result to be irreducible we thus need to show that the diagonal extension of the multimodal logic, as specified in Theorem 10, is consistent in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$. To this end we construct a DML model for the diagonal extension of the multimodal logic axiomatized by $(\Box/\blacksquare \mathit{Nec})$, $(\Box/\blacksquare K)$, $(\Box D)$ and $(\Box\blacklozenge \mathit{In})$.

Before we state our theorem we will collect some relevant information. Most importantly, we shall use the following facts. Firstly, the modal principles solely governing the modal operator ■ are true in any model based on a possible world frame. Secondly, the modal principles solely governing the modal operator □, particularly $(\Box D)$, are true in any model based on (ω, S). Lastly, $(\Box\blacklozenge \mathit{In})$ is a very simple Sahlquist formula in the sense of Blackburn et al. [2] which tells us that $(\Box\blacklozenge \mathit{In})$ defines a characteristic first-order frame property and that we can effectively compute this property.[13] We have:

Fact 26. $(\Box\blacklozenge \mathit{In})$ *defines relative to multimodal frames $\mathcal{F} = \langle W, R_{\Box}, R_{\blacksquare}\rangle$ of $\mathcal{L}_{\Box\blacksquare}$ the first order frame property*

$$(*) \qquad \forall w, v \in W(wR_{\Box}v \Rightarrow vR_{\blacksquare}w)$$

[13] See Blackburn et al. [2] for more on Sahlquist formulas and how to compute the characteristic first-order property defined by a Sahlquist formula.

Thus $(\Box\blacklozenge In)$ is true in any model based on a multimodal frame with property $(*)$.

We now state the main theorem which asserts that the diagonal extension of the multimodal logic, which Horsten and Leitgeb showed to be inconsistent in $\mathcal{L}^F_{\Box\blacksquare}$, is consistent in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$.

Theorem 27. *Let $\mathcal{S}$ be the multimodal logic in $\mathcal{L}_{\Box\blacksquare}$ axiomatized by $(\Box/\blacksquare Nec)$, $(\Box/\blacksquare K)$, $(\Box D)$ and $(\Box\blacklozenge In)$. Then $\mathcal{S}^F$ is consistent in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$.*

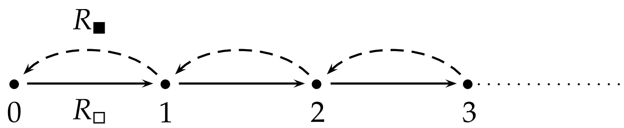
Proof. We will first construct a Kripke model of $\mathcal{L}^F_{\Box}$. We then extend the underlying frame of the model by a second accessibility relation $R_{\blacksquare}$ which will give us a model in the language $\mathcal{L}^F_{\Box} \cup \mathcal{L}_{\blacksquare}$. However, $R_{\blacksquare}$ will be converse well-founded which enables us to apply the techniques illustrated in Theorem 18 to extend the model to an appropriate model of $\mathcal{S}^F$ in $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$.

We will start by an arbitrary model M for $\mathcal{L}_{\Box}$ based on the frame (ω, S). By Theorem 21 we can expand the model to a DML model M' based on (ω, S).

In the next step we extend the frame by defining $R_{\blacksquare}$ as $R^{-1}_{\Box}$, thus

$$wR_{\Box}v \Leftrightarrow vR_{\blacksquare}w$$

With this definition $(*)$ is satisfied and therefore $(\Box\blacklozenge In)$ will be valid in the multimodal frame we have constructed. The frame can be pictured as follows:



Note that $R_{\blacksquare}$ is converse well-founded, i.e. for all nonempty $X \subseteq \omega$ there is a $w \in X$, such that for all $v \in X, \neg wR_{\blacksquare}v$.

So far we have constructed a model for the language $\mathcal{L}^F_{\Box} \cup \mathcal{L}_{\blacksquare}$ which is a model of $(\Box/\blacksquare K)$, $(\Box/\blacksquare Nec)$, $(\Box D)$ and $(\Box\blacklozenge In)$. Also the fixed-point axioms $\phi(\delta_{\phi(q)}) \leftrightarrow \delta_{\phi(q)}$ for all $\delta_{\phi(q)} \in \mathcal{L}^F_{\Box}$ are true in this model. In the last step of the model construction we will extend the model to a model of $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$ and provide a suitable interpretation for all fixed-point constants of $\mathcal{L}^F_{\blacksquare}$. As we consider $\mathcal{L}^F_{\Box} \cup \mathcal{L}^F_{\blacksquare}$, and not $\mathcal{L}^F_{\Box\blacksquare}$, we only need to provide interpretations of fixed-points of formulas in which the $\Box$-operator does not occur. This allows us to apply Smoryński's method. By Theorem 18 we know that any model of a modal language $\mathcal{L}_{\Box}$ based on a converse well-founded frame can be extended to a model of the language $\mathcal{L}^F_\Box$ in which all fixed-point axioms are true. Therefore we can extend the valuation function V' of M' to a valuation function V^F for the language $\mathcal{L}^F_\Box \cup \mathcal{L}^F_\blacksquare$, which provides a suitable interpretation of the fixed-point constants of $\mathcal{L}^F_\blacksquare$, i.e. $(\omega, S, V^F) \models \phi(\delta_{\phi(q)}) \leftrightarrow \delta_{\phi(q)}$ for all $\delta_{\phi(q)} \in \mathcal{L}^F_\blacksquare$. The resulting model establishes the consistency of the diagonal extension of the modal logic axiomatized by $(\Box/\blacksquare Nec)$, $(\Box/\blacksquare K)$, $(\Box D)$ and $(\Box\blacklozenge In)$ in $\mathcal{L}^F_\Box \cup \mathcal{L}^F_\blacksquare$. □

Corollary 28. *Horsten and Leitgeb's inconsistency result is irreducible in the sense of Definition 13.*

Corollary 29. *The inconsistency result obtained in Theorem 11 is irreducible in the sense of Definition 13.*

Proof. Direct consequence of Lemma 12. □

Consequently, we have established the existence of irreducible paradoxes in the sense of our criterion. Since we have already shown that there are also reducible paradoxes the criterion we proposed manages to differentiate between different types of paradoxes.

7 Conclusion

In this paper we discussed multimodal paradoxes which may arise when modal formulas can be diagonalized. In particular, we investigated whether these multimodal paradoxes introduce a new aspect into the discussion surrounding the modal paradoxes. The idea was that certain paradoxes arise specifically from the interaction of the different modal predicates. In this case it is not the modal properties we assumed for the different modal notions at stake that led to paradox, but the assumptions of how the modal notions interact. This distinction was made precise, for the case of two modal notions, by focusing on which instances of the diagonal lemma were required in the derivation of the paradoxes. A paradox was called a proper or irreducible paradox of interaction if diagonalizing a formula containing both modalities was required for deriving the inconsistency. If it proved sufficient to diagonalize a formula containing only one modality, we called the multimodal paradox a reducible paradox, as the paradox then seems to be nothing but a variant of a unimodal paradox.

In our paper we have established the claim that there are reducible as well as irreducible paradoxes of interaction according to the criterion we have laid out. Halbach's inconsistency result proved to be a reducible paradox upon closer inspection while Theorem 27 may be taken to confirm the claim that genuinely new paradoxes arise due to the interaction of different modalities. The theorem shows that for the "No Future" inconsistency to arise it is necessary to consider fixed-points for formulas containing both modalities, whereas fixed-points for formulas with only one modality can be consistently added. The "No Future" inconsistency discovered by Horsten and Leitgeb thus counts as an irreducible paradox, i.e. a proper paradox of interaction.

The existence of proper paradoxes of interaction suggests that multimodal paradoxes cannot be easily discovered by simply examining the underlying modal properties of some multimodal logic under consideration. Rather, whether the diagonal extension of some multimodal logic will be consistent or inconsistent depends crucially on the way the modal notions are tied together. Further investigation into, and systematization of, the proper paradoxes of interaction therefore seems to be an important topic for future research.

At this point one should mention that the present proposal has its limitations. DML can be viewed as an expressive strengthening of standard modal operator logic but clearly DML itself could be strengthened further. For example, while DML only simulates simple fixed-point equations one could strengthen the logic and allow for simultaneous fixed-point equations. In such a broader setting further paradoxes could be investigated with respect to their reducibility and irreducibility. For example, multimodal versions of the Postcard paradox could then be studied. Further extensions of DML, for example by propositional quantifiers, might also be interesting in this respect. The study of expressive strengthenings of DML should also provide insights into the robustness of our irreducibility result. If it turns out that the irreducibility of the No Future paradox can also be established in expressive strengthenings of DML this would support our result. But this study is again left for future research.

Moreover, our notion of reducibility might be only one out of many interesting notions of reducibility. For example, the "Multimodal Montague" seems to be a reducible paradox in the sense that its proof consists in rewriting Montague's original proof in the multimodal setting. That is, wherever Montague uses a single modality we replace it by two modalities and thereby arrive at the proof of Theorem 11. This suggests that nothing conceptually new happens in the case of the "Multimodal Montague", but the paradox is a proper paradox of interaction according to our definition. In following this intuition concerning the 'Multimodal Montague" one may arrive at an alternative notion of reducibility. However, we take it that our notion of reducibility is intuitively well-motivated and that it captures an important distinction. The idea of irreducible paradoxes of interaction should therefore be taken seriously.

Acknowledgements We thank Catrin Campbell-Moore and Jan Heylen for comments on drafts of this paper and an anonymous referee for a number of helpful suggestions. We also thank audiences in Munich, Paris, Leuven, Amsterdam, Buenos Aires and Irvine for helpful suggestions. Our work has been carried out within the DFG funded research project "Syntactical Treatments of Interacting Modalities". The project is hosted by the Munich Center for Mathematical Philosophy which in turn enjoys support by the Alexander von Humboldt Foundation.

References

[1] L. Alberucci and A. Facchini. On modal μ-calculus and Gödel-Löb logic. *Studia Logica*, 91(2):145–69, 2009.

[2] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.

[3] G. S. Boolos. *The Logic of Provability*. Cambridge University Press, 1993.

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

[5] V. Halbach. How not to state T-sentences. *Analysis*, 66:276–280, 2006. Correction in Analysis 67: 268.

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

[7] V. Halbach, H. Leitgeb, and P. Welch. Possible-worlds semantics for modal notions conceived as predicates. *Journal of Philosophical Logic*, 32:179–222, 2003.

[8] L. Horsten and H. Leitgeb. No future. *Journal of Philosophical Logic*, 30:259–265, 2001.

[9] D. Kaplan and R. Montague. A paradox regained. *Notre Dame Journal of Formal Logic*, 1:79–90, 1960.

[10] R. Montague. Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica*, 16:153–167, 1963.

[11] J. Myhill. Some remarks on the notion of proof. *The Journal of Philosophy*, 57:461–471, 1960.

[12] K.-G. Niebergall. Simultane objektsprachliche Axiomatisierung von Notwendigkeits- und Beweisbarkeitsprädikaten. Masterthesis, Ludwig-Maximilians Universität München, 1991.

[13] C. Smoryński. *Self-Reference and Modal Logic*. Springer Verlag, 1985.

[14] C. Smoryński. Modal logic and self-reference. In D. Gabbay and F. Guethner, editors, *Handbook of Philosophical Logic, 2nd Edition*, volume 11, pages 1–55. Kluwer Academish Publishers, 2004.

[15] T. Williamson. Iterated attitudes. In T. Smiley, editor, *Philosophical Logic. Proceedings of the British Academy 95*, pages 85–133. Oxford University Press, Oxford, 1998.