Modal Logics of Sabotage Revisited

Guillaume Aucher[1], Johan van Benthem[2], and Davide Grossi[3]

[1] University of Rennes 1,
guillaume.aucher@irisa.fr
[2] University of Amsterdam, Stanford University, Changjiang Program Tsinghua University, J.vanBenthem@uva.nl
[3] University of Liverpool,
D.Grossi@liverpool.ac.uk

Abstract. Sabotage modal logic was proposed in 2003 as a format for analyzing games that modify the graphs they are played on. We investigate some model-theoretic and proof-theoretic aspects of sabotage modal logic, which has largely come to be viewed as an early dynamic logic of graph change. Our first contribution is a characterization theorem for sabotage modal logic as a fragment of first-order logic which is invariant with respect to a natural notion of 'sabotage bisimulation'. Our second contribution is a sound and complete tableau method for analyzing reasoning in sabotage modal logic. Finally, we identify and briefly explore a number of open research problems concerning sabotage modal logic that illuminate its complexity, placing it within the current landscape of modal logics that analyze model update, and, returning to the original motivation of sabotage, fixed-point logics for network games.

1 Introduction

Sabotage modal logic (SML), first published in [35], expands the standard modal language with an edge-deletion modality $\blacklozenge\varphi$ whose intended reading is "after the deletion of at least one edge in the frame, it holds that φ". This minimal modal logic of arbitrary edge deletion stands at the beginning of a tradition of several formalisms in the dynamic-epistemic logic spirit, [36], such as 'graph modifier logic', [7], 'swap logic', [3], or 'arrow update logic', [28,1]. SML is also directly related to recent work in theoretical computer science, [32,26], learning theory, [22], logics of social networks, [33,29], and argumentation, [24,25].

Only a few properties of sabotage modal logic have been studied systematically so far. The original paper [35] showed how SML can formulate solution concepts for sabotage games, it established a Pspace upper bound for the model checking problem, and it sketched how the SML language can be translated into first-order logic, while observing that the set of validities is not closed under substitutions. The earliest more incisive results after these initial observations are in [31,30], where the multi-modal variant of SML is shown to have an undecidable satisfiability problem and a Pspace-complete model-checking problem. One main open problem mentioned in this work is a bisimulation invariance characterization for SML. This question was solved independently in the proceedings version of the present article [8] and in the work of an Argentinean team, [4]. Also independently, both teams developed tableau systems for SML to better analyze the structure of validity in this system, cf. [2,4]. We will flag and discuss some connections in the related work section at the end of this paper.[1]

Contributions and outline Our paper is structured in two parts. The first, comprising Sections 2 to 5, extends the material presented in [8] by including proofs left out of that proceedings version, while adding several new examples and a few smaller novel observations. After some preliminaries in Section 2, this first part presents two main contributions: a characterization result for SML (Section 4) as that fragment of first-order logic which is invariant for a natural notion of 'sabotage bisimulation' (Section 3) and a novel tableau system in Sections 5, 6. The second part of the article is more exploratory. Sections 7 to 10 present new material addressing the issue of what makes SML tick among current modal logics of model change, where we identify, in particular, the sources of its undecidability in a way that also sheds new light on current dynamic-epistemic logics. Moreover, we make a systematic connection with sabotage-inspired fixed-point logics for analyzing graph games, and revisit the original motivation for SML in sabotage games. Section 9 discusses related work. Section 10 concludes, showing how sabotage modal logic can serve as an interesting conceptual laboratory that generates challenging open problems of both conceptual and technical interest.

2 Preliminaries

In this section, we introduce the syntax and semantics of SML, recapitulate some key results from [30], and present a standard translation from SML into FOL.

2.1 Syntax

Let $\mathbf{P}$ be a countable set of propositional atoms. The set of formulae of the sabotage modal language $\mathcal{L}^s$ is defined by the following grammar in BNF:

$$\mathcal{L}^s : \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Diamond\varphi \mid \blacklozenge\varphi$$

where $p \in \mathbf{P}$. The remaining set of Boolean connectives $\{\vee, \rightarrow\}$ and the modal operators $\Box$ and $\blacksquare$ can be defined in the standard way. The formula $\bot$ is an abbreviation for the formula $p \wedge \neg p$ (for an arbitrarily chosen $p \in \mathbf{P}$) and $\top$ is an abbreviation for $\neg\bot$. The iteration of n sabotage operators or modalities will sometimes be denoted by $\blacklozenge^n$ and $\Diamond^n$, respectively, $\blacksquare^n$ and $\Box^n$.

To save parentheses, we rank binding strength as follows: $\blacksquare, \blacklozenge, \Box, \Diamond, \neg, \wedge, \vee, \rightarrow$. Finally, a natural measure of syntactic complexity for sabotage formulae is their 'sabotage depth', [30]. Let $\varphi \in \mathcal{L}^s$. The *sabotage depth* of φ, written $sd(\varphi)$, is inductively defined as follows: $sd(\top) = sd(p) := 0$, $sd(\neg\varphi) := sd(\varphi)$, $sd(\varphi_1 \wedge \varphi_2) := \mathtt{max}\{sd(\varphi_1), sd(\varphi_2)\}$, $sd(\Diamond\varphi) := sd(\varphi)$ and $sd(\blacklozenge\varphi) := sd(\varphi) + 1$.

[1] Actually, in [1,3,2,4], the study of SML is the start of a broader investigation of modal logics for model change, with many further themes.

2.2 Semantics

We will be working with standard relational models $\mathcal{M} = (W, R, V)$ for modal logic where W is a non-empty set; $R \subseteq W \times W$, and $V : \mathbf{P} \longrightarrow 2^W$. The pair (W, R) is called a frame, and is denoted by $\mathcal{F}$.

As usual, such semantic structures can be also interpreted as models for the binary fragment of FOL with equality[2] denoted $\mathcal{L}^1$. Sometimes we will use the following FOL terminology/notation. We say that a model $\mathcal{M}$ *satisfies* a formula $\varphi(x) \in \mathcal{L}^1$ (or a set $\Gamma(x) \subseteq \mathcal{L}^1$) with one free variable x under the assignment of w to x if and only if φ (respectively Γ) is true of w, in symbols, $\mathcal{M} \models \varphi(x)[w]$ (respectively, $\mathcal{M} \models \Gamma(x)[w]$). We say that a model $\mathcal{M}$ *realizes* a set $\Gamma(x) \subseteq \mathcal{L}^1$ with one free variable x (i.e., a type) if and only if there exists an object $w \in W$ such that $\mathcal{M} \models \Gamma(x)[w]$.

The satisfaction relation for $\mathcal{L}^s$ is defined as usual for the atomic and Boolean cases, and for the standard modalities. For the sabotage modality it is as follows:

$$(W, R, V), w \models \blacklozenge\varphi \iff \exists (w', w'') \in R \text{ s.t. } (W, R \setminus \{(w', w'')\}, V), w \models \varphi \quad (1)$$

In other words, $\blacklozenge\varphi$ is satisfied by a pointed model if and only if there exist two R-related (possibly identical) states such that, once the edge between these two states is removed from R, φ holds at the same evaluation state. The notions of validity and logical consequence are defined as usual.

We now say that two pointed models $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$ are *sabotage-related* (in formal notation, $(\mathcal{M}, w) \xrightarrow{\blacklozenge} (\mathcal{M}', w')$) if and only if: $w' = w$; $W' = W$; $R' = R \setminus \{(w'', w''')\}$ for some $w'', w''' \in W$; and $V' = V$. The set $\mathbf{r}(\mathcal{M}, w) = \{(\mathcal{M}', w') \mid (\mathcal{M}, w) \xrightarrow{\blacklozenge} (\mathcal{M}', w')\}$ consists of all models which are sabotage-related to a given pointed model $(\mathcal{M}, w)$. Similarly, the iterated variant $\mathbf{r}^n(\mathcal{M}, w) = \{(\mathcal{M}', w') \mid (\mathcal{M}_1, w_1) \xrightarrow{\blacklozenge} (\mathcal{M}_2, w_2) \xrightarrow{\blacklozenge} \ldots \xrightarrow{\blacklozenge} (\mathcal{M}_{n+1}, w_{n+1})$ & $(\mathcal{M}_1, w_1) = (\mathcal{M}, w)$ & $(\mathcal{M}_{n+1}, w_{n+1}) = (\mathcal{M}', w')\}$ denotes the set of all models which are related to $(\mathcal{M}, w)$ by a $\xrightarrow{\blacklozenge}$-path of length n. Finally, $\mathbf{r}^*(\mathcal{M}, w) = \{(\mathcal{M}', w') \mid (\mathcal{M}, w) \xrightarrow{\blacklozenge}{}^* (\mathcal{M}', w')\}$ denotes the set of all pointed models which are reachable from $(\mathcal{M}, w)$ by the reflexive and transitive closure of $\xrightarrow{\blacklozenge}$. We will often drop the reference to a given distinguished point in the model, which will be clear from the context, and simply write $\mathcal{M} \xrightarrow{\blacklozenge} \mathcal{M}'$ instead of $(\mathcal{M}, w) \xrightarrow{\blacklozenge} (\mathcal{M}', w')$.

The set of sabotage modal formulae which are satisfied by a pointed model $(\mathcal{M}, w)$, i.e., the *sabotage modal logic theory* of w in $\mathcal{M}$, is denoted $\mathbb{T}^s(\mathcal{M}, w)$. We say that two pointed models $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$ are sabotage modally equivalent —in formal notation: $(\mathcal{M}, w) \leftrightsquigarrow_s (\mathcal{M}', w')$—if and only if they satisfy the same sabotage modal formulae, that is, they have the same SML theory.

[2] One can check [12, Ch. 2.4] for basics of this first-order correspondence language.

$$\begin{tikzcd} & a & \\ & b \arrow[u] & \\ c \arrow[ur] & & d \arrow[ul] \\ & e \arrow[ul] \arrow[ur] & \end{tikzcd}$$

Fig. 1. A model showing that the general schema $\Box\varphi \to \blacksquare(\Diamond\top \to \Diamond\varphi)$ fails in SML. Let $\varphi := \Diamond\Diamond p$, let $V(p) = \{a\}$, and let the evaluation point be e.

2.3 Some notable validities and expressible properties

We list some validities of SML that demonstrate how the deletion modality works:

$$\blacksquare\bot \to \Box\bot \tag{2}$$

$$\Box\bot \to \blacksquare\Box\bot \tag{3}$$

$$p \to \blacksquare p \tag{4}$$

$$(\blacksquare p \land \neg\blacksquare\bot) \to p \tag{5}$$

$$\Box p \to \blacksquare(\Diamond\top \to \Diamond p) \tag{6}$$

$$\Diamond\varphi \land \Diamond\neg\varphi \to \blacklozenge\blacklozenge\top \tag{7}$$

where $p, q \in \mathbf{P}$. This use of atomic proposition letters is not accidental. Surprisingly many prima facie valid-looking principles fail for SML in their full schematic form with all complex substitution instances once we realize that under a deletion modality, ordinary modalities can change their truth values.

A good example is principle (6). Consider its schematic formulation

$$\Box\varphi \to \blacksquare(\Diamond\top \to \Diamond\varphi) \tag{8}$$

which states that if every accessible state satisfies φ, then after any link deletion, if the evaluation state still has a successor, it still has a φ-successor. The formula may fail if φ is modal, since deletion may happen deeper in the model and disrupt the truth of φ at successor states. See Figure 1 for an illustration.

In the above list, only the last item (7) is a schematic validity.[3] Any logic has a 'substitution core' consisting of its schematically valid principles, cf. [36], and it is an interesting open problem whether SML has a decidable, or even an axiomatizable substation core. [4]

Next, SML can express properties beyond the reach of standard modal logic. Examples are: "there are at most n successors of the current state" with $1 \leq n$

[3] Recall that a formula φ of a language $\mathcal{L}$ is a schematic validity on some class of models if $\sigma(\varphi)$ is a validity in that class for any substitution $\sigma : \mathcal{L} \longrightarrow \mathcal{L}$ of complex formulas for proposition letters.

[4] The analogous open problem for the much simpler dynamic-epistemic logic PAL of public announcement was only solved in [27]: it turned out to be decidable.

(see also Example 2 below); "there exists at least one successor or, there exists another state with at least one successor"; "there exists no successor but there is at least one state with at least one successor"; "there are exactly n edges in the model". All these properties have FOL formulations using counting quantifiers:

$$\exists^{\leq n} y\,(xRy) \tag{9}$$

$$\exists^{\leq 2} y\,(xRy) \vee \big(\exists^{\leq 1} y\,(xRy) \wedge \exists y(x \neq y \wedge \exists z(yRz)\big) \tag{10}$$

$$\neg\exists y(xRy) \wedge \exists y(\exists z(yRz)) \tag{11}$$

These properties can be expressed in SML by, in the same order:

$$\Box\bot \vee \bigvee_{1\leq i\leq n} \blacklozenge_i \Box\bot \tag{12}$$

$$\blacklozenge\Diamond\top \tag{13}$$

$$\Box\bot \wedge \blacklozenge\top \tag{14}$$

None of the above properties is expressible in the standard modal language, and defining them involves various hybrid extensions, such as *graded modalities*, [16], the *difference modality*, and the *universal modality*, [12, Ch. 7].

Another sign of strength for SML is its power to define frames up to isomorphism. For instance, it is a simple exercise to show that the formula

$$\Diamond\top \wedge \Box\Diamond\top \wedge \blacksquare\Box\bot \tag{15}$$

is true in a model if and only if its underlying frame consists of one reflexive point. This observation can be generalized to cycles of any length.

Fact 1 *For each positive natural number n, there exists an* SML*formula φ such that, for any model $\mathcal{M} = (W, R, V)$ and state s: $\mathcal{M}, s \models \varphi$ if and only if the frame (W, R) is a cycle of length n.*

Proof. We show how to build the desired formula. Define first:

$$PATH(n) \stackrel{def}{=} \bigwedge_{0\leq i\leq n} \Box_i\Diamond\top \wedge \blacksquare\left(\bigvee_{0\leq i\leq n-1} \Diamond_i\Box\bot\right)$$

Each such formula forces the existence of exactly one path of length $n+1$ from the evaluation point. The desired formula is then defined inductively as follows:

$$CYCLE(1) \stackrel{def}{=} PATH(1) = (15)$$

$$CYCLE(n+1) \stackrel{def}{=} PATH(n) \wedge \bigwedge_{1\leq i\leq n} (\neg\Diamond_{n-i}\blacklozenge_{n-i} CYCLE(i))$$

The base case is straightforward. As to the induction step, we observed above that the first conjunct forces the existence of one unique path of length at least n from the evaluation state. The second conjunct forces the $(n-1)^{th}$-step in such path to end in the evaluation state, by preventing shorter loops. It states that it is not possible to reach a state in $n-i$ steps where, after having removed these $n-i$ links, the evaluation state finds itself in a cycle of length i. □

2.4 A First-Order Translation for SML

A standard first-order translation for SML was sketched in [35] and [31]. Later on, a detailed translation was independently given in the proceedings version of this paper [8] and in [4]. In this section we describe the translation and its correctness in detail. This prepares for the later sections of the article.

Setting up the translation In order to define a translation from the language of SML into the free variable fragment of FOL with equality one needs to keep track of the changes that the sabotage operators introduce in the model.

This can be achieved by indexing the standard translation with a set E consisting of pairs of variables. The idea is that when the standard translation is applied to the outermost operator of a given formula, this set is empty. As the analysis proceeds towards inner operators, each sabotage operator $\blacklozenge$ in the formula will introduce a new pair of variables in E, which will be bound by an existential quantifier. Here is the formal definition:

Definition 1 (Standard translation for SML). *Let E be a set of pairs (y,z) of variables—standing for edges—and let x be a designated variable. The translation $ST_x^E : \mathcal{L}^s \longrightarrow \mathcal{L}^1$ is recursively defined as follows:*

$$
\begin{aligned}
ST_x^E(p) &= P(x)\\
ST_x^E(\bot) &= x \neq x\\
ST_x^E(\neg\varphi) &= \neg ST_x^E(\varphi)\\
ST_x^E(\varphi_1 \wedge \varphi_2) &= ST_x^E(\varphi_1) \wedge ST_x^E(\varphi_2)\\
ST_x^E(\Diamond\varphi) &= \exists y \left(xRy \wedge \bigwedge_{(v,w)\in E} \neg(x = v \wedge y = w) \wedge ST_y^E(\varphi) \right)\\
ST_x^E(\blacklozenge\varphi) &= \exists y,z \left(yRz \wedge \bigwedge_{(v,w)\in E} \neg(y = v \wedge z = w) \wedge ST_x^{E\cup\{(y,z)\}}(\varphi) \right)
\end{aligned}
$$

The key inductive clauses here obviously concern $\Diamond$-formulae and $\blacklozenge$-formulae. Formula $\blacklozenge\varphi$ is translated as the first-order formula stating that there exists some R-edge denoted by (y,z); that such an edge is different from any edge already denoted by the pairs in E; that the translation of φ should now be carried out with respect to the set $E \cup \{(y,z)\}$; and that this translation is realized at x.

The standard modality $\Diamond\varphi$ translates as a first-order formula, with x free, which states the existence of a state y accessible from x via an edge different from all the edges in the set E, and that the translation of φ is realized at y.

Setting up the translation like this allows one to book-keep the removal of edges via E. The removal of edges is handled by imposing the existence of states which are different from the ones reachable via the 'removed' edges. In other words edge removal is simulated by imposing the existence of edges which are then not used to interpret inner modal operators.

It is important to notice the following feature of the translation. Depending on the chosen E, ST^E can possibly yield formulae with several free variables, e.g.: $ST_x^{(v,w)}\Diamond p = \exists y\,(xRy \wedge \neg(x = v \wedge y = w) \wedge p)$. However, if ST^E is applied to a formula φ by setting $E = \emptyset$, that is to say, if the translation is initiated with an empty E, then, at each successive application of ST^E to subformulae of φ, the variables occurring in W will be bound by some quantifiers introduced at previous steps. For any φ, $ST_x^{\emptyset}(\varphi)$ yields a FOL formula with only x free.

Example 1. We illustrate the translation by means of an example:

$$\begin{aligned}
ST_x^{\emptyset}(\blacklozenge\Diamond\blacklozenge\Diamond p) &= \exists y, z\left(yRz \wedge ST_x^{\{(y,z)\}}(\Diamond\blacklozenge\Diamond p)\right)\\
&= \exists y, z\left(yRz \wedge \exists v\left(xRv \wedge \neg(x = y \wedge v = z) \wedge ST_v^{\{(y,z)\}}(\blacklozenge\Diamond p)\right)\right)\\
&= \exists y\,(xRy \wedge \exists v\,(xRv \wedge \neg(x = x \wedge y = z)\wedge\\
&\quad\ \exists y', z'\left(y'Rz' \wedge \neg(y' = y \wedge z' = z) \wedge ST_v^{\{(y,z),(y',z')\}}(\Diamond p)\right)\right)\Big)\\
&= \exists y, z\,(yRz \wedge \exists v\,(xRv \wedge \neg(x = x \wedge y = z)\wedge\\
&\quad\ \exists y', z'\,(y'Rz' \wedge \neg(y' = y \wedge z' = z)\wedge\\
&\quad\ \exists v'\,(vRv' \wedge \neg(v = y \wedge v' = z) \wedge \neg(v = y' \wedge v' = z')\wedge\\
&\quad\ ST_{v'}^{\{(y,z),(y',z')\}}(p)))))\\
&= \exists y, z\,(yRz \wedge \exists v\,(xRv \wedge \neg(x = x \wedge y = z)\wedge\\
&\quad\ \exists y', z'\,(y'Rz' \wedge \neg(y' = y \wedge z' = z)\wedge\\
&\quad\ \exists v'\,(vRv' \wedge \neg(v = y \wedge v' = z) \wedge \neg(v = y' \wedge v' = z') \wedge p))))
\end{aligned}$$

This formula states that after some sabotage it is still possible to reach a state where, after a second sabotage, a p-state can be reached.

Correctness of the translation To check this complex syntax, we now sketch a proof of the correctness of the translation proposed in Definition 1.

Theorem 1. *Let $\mathcal{M}, w$ be a pointed model and $\varphi \in \mathcal{L}^s$:*

$$\mathcal{M}, w \models \varphi \iff \mathcal{M} \models ST_x^{\emptyset}(\varphi)[w]$$

Proof. This goes by induction on the structure of φ. The key inductive step for the sabotage operator $\blacklozenge$ is proven by the following series of equivalences:

$$\begin{aligned}
\mathcal{M}, w \models \blacklozenge\varphi &\iff \mathcal{M}, w \xrightarrow{\blacklozenge} \mathcal{M}', w \models \varphi && \text{semantics of } \blacklozenge\ (1)\\
&\iff \mathcal{M}, w \xrightarrow{\blacklozenge} \mathcal{M}' \models ST_x^{\emptyset}(\varphi)[w] && \text{Inductive Hypothesis}\\
&\iff \mathcal{M} \models \exists y, z\left(yRz \wedge ST_x^{\{(y,z)\}}(\varphi)[w]\right) && \text{semantics } \blacklozenge\ (1) + \text{Def. } 1\\
&\iff \mathcal{M} \models ST_x^{\emptyset}(\blacklozenge\varphi)[w] && \text{Def. } 1 \quad \square
\end{aligned}$$

While this translation is mainly a tool for us, it does exhibit some interesting features. The standard translation for modal logic takes formulas into the two-variable fragment of first-order logic. Here, however, things are different.

Proposition 1. SML *is not contained in any fixed variable fragment of* FOL.

Proof. Consider the SML formulae of Section 2 with counting quantifiers for at most n accessible worlds. It is well-known that no fixed finite-variable fragment $FO(k)$ of FOL can define all of these, since the Ehrenfeucht k-pebble game that is characteristic for such a fragment, cf. [37, Ch. 14], cannot distinguish between pointed models having at most k and at most $k+1$ accessible worlds. □

Finally, it may be pointed out that our translation can be viewed in two ways. It may be seen as showing that SML formulas are pretty messy first-order formulas, involving a lot of variable binding. But it may also be seen as showing that SML is a quite succinct variable-free notation for a complex part of FOL. [5]

3 Bisimulation for SML

In this section, we introduce a notion of 'sabotage bisimulation' for SML, a task left open in [31,30]. The invariance results enabled by this new notion of bisimulation (that is, Propositions 2 and 3 below) were independently proven in [8] and [4], albeit in slightly different formulations.

3.1 Sabotage Bisimulation

Definition 2 (s-bisimulation). *Let $\mathcal{M}_1 = (W_1, R_1, V_1)$ and $\mathcal{M}_2 = (W_2, R_2, V_2)$ be two relational models. A non-empty relation $Z \subseteq \mathbf{r}^*(\mathcal{M}_1, w) \times \mathbf{r}^*(\mathcal{M}_2, v)$ is an* s-bisimulation *between the two pointed models $(\mathcal{M}_1, w)$ and $(\mathcal{M}_2, v)$—notation, $Z : (\mathcal{M}_1, w) \leftrightarrows_s (\mathcal{M}_2, v)$—if the following conditions are satisfied:*

Atom: *If $(\mathcal{M}_1, w)Z(\mathcal{M}_2, v)$ then $\mathcal{M}_1, w \models p$ iff $\mathcal{M}_2, v \models p$, for any atom p.*
Zig$_\Diamond$: *If $(\mathcal{M}_1, w)Z(\mathcal{M}_2, v)$ and there exists $w' \in W_1$ s.t. wR_1w' then there exists $v' \in W_2$ s.t. vR_2v' and $(\mathcal{M}_1, w')Z(\mathcal{M}_2, v')$;*
Zag$_\Diamond$: *If $(\mathcal{M}_1, w)Z(\mathcal{M}_2, v)$ and there exists $v' \in S_s$ s.t. vR_1v' then there exists $w' \in W_1$ s.t. wR_1w' and $(\mathcal{M}_1, w')Z(\mathcal{M}_2, v')$;*
Zig$_\blacklozenge$: *If $(\mathcal{M}_1, w)Z(\mathcal{M}_2, v)$ and there exists $\mathcal{M}'_1$ such that $(\mathcal{M}_1, w) \xrightarrow{\blacklozenge} (\mathcal{M}'_1, w)$, then there exists $\mathcal{M}'_2$ such that $(\mathcal{M}_2, v) \xrightarrow{\blacklozenge} (\mathcal{M}'_2, v)$ and $(\mathcal{M}'_1, w)Z(\mathcal{M}'_2, v)$;*
Zag$_\blacklozenge$: *If $(\mathcal{M}_1, w)Z(\mathcal{M}_2, v)$ and there exists $\mathcal{M}'_2$ such that $(\mathcal{M}_2, v) \xrightarrow{\blacklozenge} (\mathcal{M}'_2, v)$, then there exists $\mathcal{M}'_1$ such that $(\mathcal{M}_1, w) \xrightarrow{\blacklozenge} (\mathcal{M}'_1, w)$ and $(\mathcal{M}'_1, w)Z(\mathcal{M}'_2, v)$.*

We write $(\mathcal{M}_1, w) \leftrightarrows_s (\mathcal{M}_2, v)$ if there is an s-bisimulation Z s.t. $(\mathcal{M}_1, w)Z(\mathcal{M}_2, v)$.

The notion of s-bisimulation strengthens the standard modal bisimulation with back and forth conditions for the sabotage modality. Here, just as the sabotage modality is an 'external' modality looking across different models, so is s-bisimulation an 'external' notion of bisimulation. Standard bisimulation keeps a relational graph model fixed and changes the evaluation point along the accessibility relation of the model, s-bisimulation keeps the evaluation point fixed but changes the model by picking one among the sabotage-accessible ones.

[5] Stating and proving precise succinctness results for SML is in fact one more interesting open problem about sabotage modal logic.

3.2 Bisimulation and Modal Equivalence in SML

We first show that s-bisimulation implies SML equivalence.

Proposition 2 ($\leftrightarroweq_s \subseteq \leftrightsquigarrow_s$)**.** *For any two pointed models $(\mathcal{M}_1, w)$ and $(\mathcal{M}_2, v)$, if $(\mathcal{M}_1, w) \leftrightarroweq_s (\mathcal{M}_2, v)$, then $(\mathcal{M}_1, w) \leftrightsquigarrow_s (\mathcal{M}_2, v)$.*

Proof. The proof is by induction on the syntax of φ. Assume $(\mathcal{M}_1, w_1)Z(\mathcal{M}_2, w_2)$. *Base:* The Atom clause of Definition 2 covers the case of atoms and propositional constants. *Induction Step:* The Boolean cases are routine as usual. The Zig$_\Diamond$ and Zag$_\Diamond$ clauses of Definition 2 take care of $\Diamond$-formulae in a standard way familiar from basic modal logic. As for $\blacklozenge$-formulae, assume that $\mathcal{M}_1, w_1 \models \blacklozenge\varphi$. By the semantics of $\blacklozenge$, we have $\mathcal{M}_1 \xrightarrow{\blacklozenge} \mathcal{M}'_1, w \models \varphi$ and, by clause Zig$_\blacklozenge$ of Definition 2, it follows that $\mathcal{M}_2 \xrightarrow{\blacklozenge} \mathcal{M}'_2$ and $(\mathcal{M}'_1, w)Z(\mathcal{M}'_2, v)$. By the inductive hypothesis, we conclude that $\mathcal{M}'_2, v \models \varphi$ and, consequently, $\mathcal{M}_2, v \models \blacklozenge\varphi$. Similarly, from $\mathcal{M}_2, v \models \blacklozenge\varphi$, we conclude $\mathcal{M}_1, w \models \blacklozenge\varphi$ by clause Zag$_\blacklozenge$ of Definition 2. □

Just as for the standard modal language, the converse of Proposition 2 can be proven under the assumption that the models at issue are 'ω-saturated'. Before introducing this notion let us fix some notation. Given a finite set Y, the expansion of $\mathcal{L}^1$ with a finite set of constants Y is denoted by $\mathcal{L}^1_Y$, and the expansion of a relational model $\mathcal{M}$ to $\mathcal{L}^1_Y$ is denoted by $\mathcal{M}_Y$.[6]

Definition 3 (**ω-saturation**)**.** *A model $\mathcal{M} = (W, R, V)$ is ω-saturated if, for every $Y \subseteq W$ such that $|Y| < \omega$, the expansion $\mathcal{M}_Y$ realizes every set $\Gamma(x)$ of $\mathcal{L}^1_Y$-formulae whose finite subsets $\Gamma'(x) \subseteq \Gamma(x)$ are all realized in $\mathcal{M}_Y$.*

Spelled out a bit further, a model $\mathcal{M}$ is *ω-saturated* if for any set of formulae $\Gamma(x, y_1, \ldots, y_n)$ over a finite set of variables, once some interpretation of $y_1, \ldots, y_n$ is fixed to, say, $w_1, \ldots, w_n$, and all finite subsets of $\Gamma(x)[w_1, \ldots, w_n]$ are realizable in $\mathcal{M}$, then the whole of $\Gamma(x)[w_1, \ldots, w_n]$ is realizable in $\mathcal{M}$. From a modal point of view, Definition 3 requires that, if for any subset of Γ there are accessible states satisfying it at the evaluation point, then there are accessible states satisfying the whole of Γ at the evaluation point. This is precisely the property used in the proof of the following proposition.

Proposition 3 ($\leftrightsquigarrow_s \subseteq \leftrightarroweq_s$)**.** *For any two ω-saturated pointed models $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$, if $(\mathcal{M}_1, w_1) \leftrightsquigarrow_s (\mathcal{M}_2, w_2)$, then $(\mathcal{M}_1, w_1) \leftrightarroweq_s (\mathcal{M}_2, w_2)$.*

Proof. It suffices to show that the relation $\leftrightsquigarrow_s$ itself is an s-bisimulation in the sense of Definition 2. *Base Case:* The condition Atom is straightforwardly satisfied, being a special case of modal equivalence of points. *Back and Forth Conditions:* The proof for conditions Zig$_\Diamond$ and Zag$_\Diamond$ proceeds as usual for basic modal languages. We prove that the condition Zig$_\blacklozenge$ is satisfied. Assume that $(\mathcal{M}_1, w_1) \leftrightsquigarrow_s (\mathcal{M}_2, w_2)$ and $(\mathcal{M}_1, w_1) \xrightarrow{\blacklozenge} (\mathcal{M}'_1, w_1)$. We show there must be a model $(\mathcal{M}'_2, w_2)$ such that $(\mathcal{M}_2, w_2) \xrightarrow{\blacklozenge} (\mathcal{M}'_2, w_2)$ and $(\mathcal{M}'_1, w_1) \leftrightsquigarrow_s (\mathcal{M}'_2, w_2)$.

[6] For more on ω-saturation we refer the reader to [12, Ch. 2] and [13, Ch. 2].

For a start, we have that for any finite $\Gamma \subseteq \mathbb{T}^s(\mathcal{M}'_1, w_1)$ the following sequence of equivalences holds:

$$\begin{aligned}
\mathcal{M}_1, w_1 \models \blacklozenge \bigwedge \Gamma &\iff \mathcal{M}_2, w_2 \models \blacklozenge \bigwedge \Gamma \\
&\iff \mathcal{M}_2 \models ST_x^{\emptyset}\left(\blacklozenge \bigwedge \Gamma\right)[w_2] \\
&\iff \mathcal{M}_2 \models \exists y, z \left(yRz \wedge ST_x^{\{(y,z)\}}\left(\bigwedge \Gamma\right)\right)[w_2]
\end{aligned}$$

The first equivalence holds by the assumption of sabotage equivalence between $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$. The second one follows by Theorem 1 and the third one by Definition 1. From this, by ω-saturation of $\mathcal{M}_2$ we can conclude that:

there are $y, z \in \mathcal{M}_2$ such that yRz and $\mathcal{M}_2 \models ST_x^{\{(y,z)\}}(\mathbb{T}^s(\mathcal{M}'_1, w_1))[w_2]$.

By Theorem 1 there exists then a model $\mathcal{M}'_2$ such that $\mathcal{M}_2 \xrightarrow{\blacklozenge} \mathcal{M}'_2$ and $\mathcal{M}'_2 \models ST_x^{\emptyset}(\mathbb{T}^s(\mathcal{M}'_1, w_1))[w_2]$. By Theorem 1 we conclude that $(\mathcal{M}'_1, w_1) \leftrightsquigarrow_s (\mathcal{M}'_2, w_2)$, which completes the proof of the Zig$_\blacklozenge$ clause.

In the same way it can be shown that also the condition Zag$_\blacklozenge$ is satisfied. □

We have thus established a precise match between sabotage modal equivalence and sabotage bisimulation for the special class of ω-saturated models. [7] Now we can move to our main intended result.

4 Characterization of SML by Invariance

In this section, we obtain an analogue of one more classical result for basic modal logic. We characterize SML as the one free variable fragment of FOL which is invariant under s-bisimulation.[8]

Theorem 2 (Characterization of SML by s-bisimulation invariance). *An $\mathcal{L}_1$-formula is equivalent to the translation of an $\mathcal{L}^s$ formula if, and only if, it is invariant for sabotage bisimulation.*

Proof. The direction from left to right follows from Proposition 2. In the opposite direction, we proceed as customary. Let $\varphi \in \mathcal{L}^1$ with one free variable x. Assume that φ is invariant under s-bisimulation and consider the following set:

$$\mathbb{C}^s(\varphi) = \{ST_x^{\emptyset}(\psi) \mid \psi \in \mathcal{L}^s \text{ and } \varphi \models ST_x^{\emptyset}(\psi)\}.$$

The result is a direct consequence of the following two claims:

[7] A way of obtaining this result for any two models, whether saturated or not, is by using an *infinitary* version of SML, whose details we forego here.

[8] Recall that the standard translation $ST^{\emptyset}$ of a sabotage modal logic formula always produces a FOL formula with only one free variable.

(a) If $\mathbb{C}^s(\varphi) \models \varphi$, then φ is equivalent to the translation of an $\mathcal{L}^s$-formula.
(b) In fact, $\mathbb{C}^s(\varphi) \models \varphi$ – that is, for any pointed model $\mathcal{M}, w$: if $\mathcal{M} \models \mathbb{C}^s(\varphi)[w]$, then $\mathcal{M} \models \varphi[w]$.

Claim (a). Assume that $\mathbb{C}^s(\varphi) \models \varphi$. From the deduction and compactness theorems for FOL, we have that $\models \bigwedge \Gamma \to \varphi$ for some finite $\Gamma \subset \mathbb{C}^s(\varphi)$. The converse holds by the definition of $\mathbb{C}^s(\varphi)$: $\models \varphi \to \bigwedge \Gamma$. We thus have that $\models \varphi \leftrightarrow \bigwedge \Gamma$, proving the claim.

Claim (b). Take any pointed model $\mathcal{M}, w$ such that $\mathcal{M} \models \mathbb{C}^s(\varphi)[w]$ and consider its sabotage modal theory $\mathbb{T}^s(\mathcal{M}, w)$. Now consider the set of formulas $\Sigma = ST_x^{\emptyset}(\mathbb{T}^s(\mathcal{M}, w)) \cup \{\varphi\}$. We first show that:

(c) Σ is consistent.

To prove (c), assume, towards a contradiction, that Σ is inconsistent. By the compactness of FOL we then obtain that $\models \varphi \to \neg \bigwedge \Gamma$ for some finite $\Gamma \in \Sigma$. But then, by the definition of $\mathbb{C}^s(\varphi)$, we have that $\neg \bigwedge \Gamma \in \mathbb{C}^s(\varphi)$, and hence $\neg \bigwedge \Gamma \in ST_x^{\emptyset}(\mathbb{T}^s(\mathcal{M}, w))$, which is impossible since $\Gamma \subset ST_x^{\emptyset}(\mathbb{T}^s(\mathcal{M}, w))$.

Now Claim (b) follows if we can show that

(d) $\mathcal{M} \models \varphi[w]$.

Here is a proof for (d). As Σ is consistent, it can be satisfied by a pointed model, say, $\mathcal{M}', w'$. Observe, first of all, that $\mathcal{M}, w \leftrightsquigarrow_s \mathcal{M}', w'$ as they both have the same sabotage modal theory. Now take two ω-saturated elementary extensions $(\mathcal{M}_\omega, w)$ and $(\mathcal{M}'_\omega, w')$ of $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$. That such extensions exist can be proven by a standard chain construction argument, [13, Proposition 3.2.6].

Then, by the invariance of FOL under elementary extensions, since $\mathcal{M}' \models \varphi[w]$ (by the construction of Σ), we can conclude that $\mathcal{M}'_\omega \models \varphi[w]$. From this, by the assumption that φ is invariant for s-bisimulation and Proposition 3, we conclude that $\mathcal{M}_\omega \models \varphi(x)[w]$ – and again, by elementary extension, that $\mathcal{M} \models \varphi(x)[w]$, and this completes the proof. $\square$

Definable and undefinable properties in SML. So which FOL properties belong to the fragment identified by Theorem 2 and which ones do not? We provide examples of SML-definable and undefinable properties of models.

Example 2 (Counting successors). Consider the earlier-discussed FOL property "there exist at most n successors" of the current point. This property is not standard modal bisimulation invariant, but it is easy to see that it is invariant with respect to sabotage bisimulation. It is therefore definable in SML, something that we had established already by a direct argument.

Example 3 (Confluence). Consider the FOL property "every two successors of the current point have a shared successor". This property is not invariant for sabotage bisimulation, witness Figure 2, and hence it is not definable in SML.

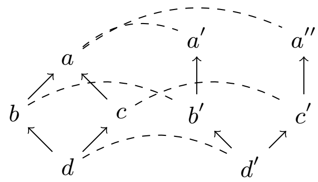

Fig. 2. Two s-bisimilar models (the s-bisimulation runs via the dashed lines). At state d to the left, the property "every two successors share a joint successor" is true. But it fails at point d' in the model to the right.

Technically, to draw a sabotage bisimulation, one should draw not only the two base models, but also all sub-models obtainable by one edge deletion, and so forth. So it is worth spending some time to see why the dashed lines in Figure 2 indeed form a sabotage bisimulation Z between the model $\mathcal{M}$ on the left and $\mathcal{M}'$ on the right. Notice that the accessibility relations in both models have the same cardinality, that is, 4. We proceed inductively as follows.[9] Consider $\mathbf{r}^i(\mathcal{M}, d)$ and $\mathbf{r}^i(\mathcal{M}', d')$ with $1 \leq i \leq 4$, that is the set of submodels that can be reached on the left, respectively right, model via exactly i edge deletions. For the base case: for $\mathcal{M}_4, s \in \mathbf{r}^4(\mathcal{M}, d)$ and $\mathcal{M}'_4, s' \in \mathbf{r}^4(\mathcal{M}', d')$, it is obvious that $(\mathcal{M}_4, s)Z(\mathcal{M}'_4, s')$ is a sabotage bisimulation for any s, s' with sZs' as in the picture, since no successors exist and no further deletion can be carried out. For the induction step, first notice that, for any $1 \leq i \leq 4$, for any $\mathcal{M}_{i-1} \in \mathbf{r}^{i-1}(\mathcal{M}, d)$, there exists $\mathcal{M}'_{i-1} \in \mathbf{r}^{i-1}(\mathcal{M}', d')$ such that $(\mathcal{M}_{i-1}, s)Z(\mathcal{M}'_{i-1}, s')$, with sZs' as in the picture, is a standard bisimulation. That is, for any pointed submodel reachable on the left with i deletions, there exists a pointed submodel on the right for which Z is a standard modal bisimulation. Then, to complete the induction step, we need to show that, if $(\mathcal{M}_i, s)Z(\mathcal{M}'_i, s')$ is a sabotage bisimulation, with sZs' as in the picture, then for any $\mathcal{M}_{i-1} \in \mathbf{r}^{i-1}(\mathcal{M}, d)$, there exists $\mathcal{M}'_{i-1} \in \mathbf{r}^{i-1}(\mathcal{M}', d')$ such that $(\mathcal{M}_{i-1}, s)Z(\mathcal{M}'_{i-1}, s')$ is a sabotage bisimulation, and vice versa. In words, if two pointed submodels reachable via i deletions are sabotage bisimilar, then for any pointed submodel reachable via $i-1$ deletions on the left, there exists a pointed submodel reachable via $i-1$ deletions on the right such that Z is a sabotage bisimulation connecting the two. By what was established above, a visual inspection of Z shows that this is indeed the case.

Example 4 (Reflexive states). Consider the FOL property xRx. This property is not invariant with respect to sabotage bisimulation. To witness this fact take two models $\mathcal{M} = (\mathbb{N}, \geq)$ and $\mathcal{M}' = (\mathbb{N}, >)$ on the set of natural numbers (with 0 as the distinguished point in each case) where the accessibility relations are: (a) on the first model, the 'greater or equal' relation (reflexive), and (b) on the second model, the strictly greater relation (irreflexive).

Now we have that $(\mathcal{M}, 0) \leftrightarroweq_s (\mathcal{M}', 0)$. Figure 3 shows this fact by depicting (part of) a relation which is a standard modal bisimulation Z between the two

[9] Technically speaking this reasoning involves a double induction.

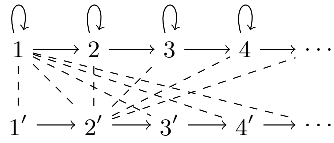

Fig. 3. Two infinite chains $(\mathbb{N}, \geq)$ and $(\mathbb{N}, >)$, with transitive edges omitted in both frames. Only the part of the s-bisimulation relation originating in points 1 and $2'$ is depicted. The remaining edges at other points are positioned similarly.

models, but which in addition has the property that any edge deletion on one model can be 'mirrored' on the other model to obtain pointed models that are still connected by Z (in the sense of Definition 2). In particular, in the picture, observe that deletion of a reflexive edge in $\mathcal{M}$ at point i can be 'mirrored' by the deletion of edge $(i, i+1)$ in $\mathcal{M}'$ (here, note that the accessibility relations are transitive in both models). However, $\mathcal{M} \models xRx[0]$ and $\mathcal{M}' \not\models xRx[0]$. The reflexivity property xRx is therefore not definable in SML.

5 A Tableau Method for SML

What about explicit proof calculi for reasoning in sabotage modal logic? As we shall see later, unlike dynamic-epistemic logics, SML does not support recursion axioms that can be added in a straightforward manner to a standard modal base logic to axiomatize the complete dynamic system. Indeed, we have not been able to find a natural Hilbert-style formulation for SML. Therefore, we turn to a method that is easier to formulate staying close to the semantic meaning of searching for counter-examples. This flexibility is provided by *semantic tableaux.*

Tableaux for SML cannot be quite like their standard modal counterparts for systems like K (cf. [18]), since there are some obvious background differences. For instance, SML does not have the tree-model property: there are specific SML formulae satisfied in relational models whose underlying frames can *not* be trees. [10] The following procedure is the adaptation that we need.

Definition 4 (Label, labeled formula and relation term). *Let S be an infinite set whose elements are called* labels. *An* extended label *is an expression of the form ℓ^E where $\ell \in S$ and E is a finite set of pairs of S. A* labeled formula *is an expression of the form $(\ell^E\ \varphi)$ where ℓ^E is a label and $\varphi \in \mathcal{L}^s$. A* relation term *is an expression of the form $(R\ \ell_1\ \ell_2)$ where $\ell_1, \ell_2 \in S$.*

Labels ℓ represent possible worlds of modal models, while a relation term $(R\ \ell_1\ \ell_2)$ represents that the pair of worlds represented by (ℓ_1, ℓ_2) belongs to the accessibility relation R. The set of pairs E in ℓ^E represents the pairs of worlds

[10] For example, our earlier formula $\Diamond\top \wedge \Box\Diamond\top \wedge \blacksquare\Box\bot$ in Section 2 was true in a model if and only if its underlying frame consists of one reflexive point.

Input: A formula $\varphi \in \mathcal{L}^s$.

Output: A tableau $\mathcal{T}$ for φ: each branch may be infinite, finite and labeled open, or finite and labeled closed.

1. Initially, $\mathcal{T}$ is a tree consisting of a single root node labeled with $(\ell^{\emptyset}\ \varphi)$.
2. Repeat the following steps as long as possible:
   (a) *Choose* a branch which is neither closed nor open and choose a labeled formula $(\ell^E\ \psi)$ (or a pair of labeled formula $(\ell^E\ \psi)$ and relation term $(R\ \ell_1\ \ell_2)$) not selected before on this branch.
   (b) *Apply* the appropriate tableau rule of Figure 5 to $(\ell^E\ \psi)$ (or the pair $(\ell^E\ \psi)$, $(R\ \ell_1\ \ell_2)$):
   - if the tableau rule is rule $\neg\wedge$ (or rules $\Diamond$, $\blacklozenge$), add two successor nodes (resp. $n+1$, n successor nodes) to the branch labeled with the instantiations of the denominators of that rule,
   - otherwise, add a unique successor node labeled with the instantiation of the denominator(s) of that rule.

   (c) i. *Label* by $\times$ (*closed*) the (new) branches which contain two labeled formulae $(\ell^E\ p)$ and $(\ell^F\ \neg p)$ (where E and F may possibly be different sets) or two labeled formulae $(\ell^E\ \varphi)$ and $(\ell^E\ \neg\varphi)$.
   ii. *Label* by $\odot$ (*open*) the (new) branches where there are no more formulae to decompose.

Fig. 4. Construction of an SML tableau.

of the accessibility relation R that have been removed by application of the rule for $\blacklozenge$ from the Kripke model constructed at this stage by the tableau method.

Definition 5 (Tableau). *A* (labeled) tableau *is a tree whose nodes are labeled with labeled formulae or relation terms. The tableau tree for a formula is constructed as shown in the algorithm of Figure 4. In the tableau rules of Figure 5, the formulae above the horizontal lines are called* numerators *and those below are called* denominators. *A tableau* closes *when all its branches are closed. A branch is* open *when it is infinite or it terminates in a leaf labeled open.*

We briefly discuss the modal and sabotage rules, those for the propositional connectives being as usual. The rules $\neg\Diamond$ and $\neg\blacklozenge$ are natural adaptations of the standard rule $\neg\Diamond$ of modal logic. The only difference is that they can be applied only to relation terms representing edges not already removed by the sabotage modality, and therefore not present in E. For the rule $\neg\blacklozenge$, moreover, the edge removed is added to the set E of the edges already removed. The rule $\Diamond$ is more complex than the standard modal rule, because SML does not have the tree-model property, unlike standard modal logic. So, we have to consider not only that a new world/label ℓ_{n+1} accessible from ℓ may satisfy φ, but also that one of the worlds/labels ℓ_i of the current model which have already been introduced may satisfy φ – and in that case, we also put an accessibility relation from ℓ to this old world/label ℓ_i. The rule $\blacklozenge$ follows the same kind of resasoning: we may

$$\frac{(\ell^E\ \varphi \wedge \psi)}{(\ell^E\ \varphi)\ (\ell^E\ \psi)}\ \wedge \qquad \frac{(\ell^E\ \neg(\varphi \wedge \psi))}{(\ell^E\ \neg\varphi) \mid (\ell^E\ \neg\psi)}\ \neg\wedge \qquad \frac{(\ell^E\ \neg\neg\varphi)}{(\ell^E\ \varphi)}\ \neg\neg$$

$$\frac{(\ell_1^E\ \neg\Diamond\varphi) \quad (R\ \ell_1\ \ell_2)}{(\ell_2^E\ \neg\varphi)}\ \neg\Diamond \qquad \frac{(\ell^E\ \neg\blacklozenge\varphi) \quad (R\ \ell_1\ \ell_2)}{(\ell^{E\cup\{(\ell_1,\ell_2)\}}\ \neg\varphi)}\ \neg\blacklozenge$$

where $(\ell_1, \ell_2) \notin E$ in both rules above.

$$\frac{(\ell^E\ \Diamond\varphi)}{(R\ \ell\ \ell_1)(\ell_1^E\ \varphi) \mid \ \dots \ \mid (R\ \ell\ \ell_n)(\ell_n^E\ \varphi) \mid (R\ \ell\ \ell_{n+1})(\ell_{n+1}^E\ \varphi)}\ \Diamond$$

where $\{\ell_1, \dots, \ell_n\}$ are all the labels occurring in the current branch such that $(\ell, \ell_i) \notin E$ for all $i \in \{1, \dots, n\}$ and ℓ_{n+1} is a 'fresh' label not occurring in the current branch.

$$\frac{(\ell^E\ \blacklozenge\varphi)}{(R\ \ell_1\ \ell_1')(\ell^{E\cup\{(\ell_1,\ell_1')\}}\ \varphi) \mid \ \dots \ \mid (R\ \ell_n\ \ell_n')(\ell^{E\cup\{(\ell_n,\ell_n')\}}\ \varphi)}\ \blacklozenge$$

where $\{(\ell_1, \ell_1'), \dots, (\ell_n, \ell_n')\} := (M \times M) \cup \{(\ell_+, \ell_{++})\} \setminus E$, with M the set of labels occurring in the current branch to which we add a 'fresh' label ℓ_*, and (ℓ_+, ℓ_{++}) is a pair of 'fresh' and distinct labels.

Fig. 5. Tableau rules.

remove an edge represented by a pair of worlds/labels which has already been introduced or which contains either one or two 'fresh' worlds/labels.

The construction of a tableau need not terminate: see Example 5. This is in line with the fact that the satisfiability problem of SML is undecidable. [11]

Example 5. In Figure 6, on the right, we display the execution of the tableau method of Figure 4 on the formula $\Diamond\top \wedge \Box\Diamond\top \wedge \blacksquare\Box\bot$. We obtain a single open branch (labeled with $\odot$) from which we can extract a model whose frame is a single reflexive point. This formula is thus satisfiable, and in fact only in this frame. In Figure 6, on the left, we show that the tableau construction may not necessarily terminate by exhibiting an infinite branch in the tableau for the formula $\Diamond\top \wedge \blacksquare\Diamond\top$. Even if the formula holds in finite pointed models having at least two successors, our tableau method does not terminate with this formula as input, and it produces a pointed model with infinitely many successors.

Finally, the reader may find it of interest to create a tableau for a formula that really requires infinite models. Instead of doing this in tableau format, we end by noting a particularly easy case where the Finite Model Property fails.

[11] If we remove the rules for sabotage we get a sound and complete tableau method for logic K that is somewhat non-standard and computationally demanding.

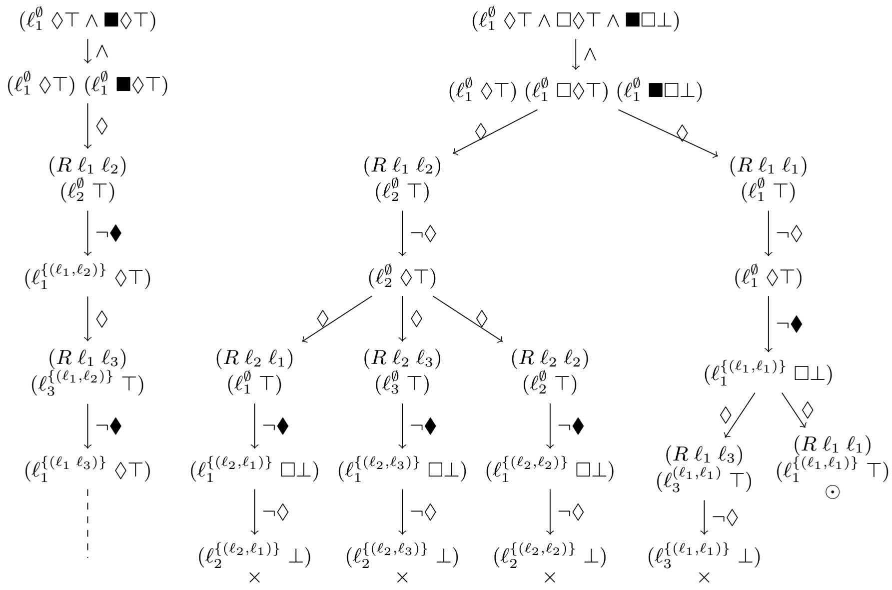

Fig. 6. Two extreme scenarios. An infinite branch arising in the tableau for the formula $\Diamond\top \wedge \blacksquare\Diamond\top$ (*left*). A finite tableau for $\Diamond\top \wedge \Box\Diamond\top \wedge \blacksquare\Box\bot$ (*right*).

Example 6 (Failure of FMP in an extended language). A somewhat complex formula for the failure of FMP is given in [17]. Here we give a more concise formula that, however, uses some extra expressivity resources. Let us extend the SML language a little by including a converse modality P (for 'past'—while we can think of the above $\Diamond$ as 'future') plus a universal modality U over all worlds. The successive conjuncts of the following formula then say that the accessibility relation is a function, that is one-to-one, but not surjective:

$$U((\Diamond\top \wedge \blacklozenge\Box\bot) \wedge (P\top \rightarrow \blacklozenge\neg P\top) \wedge E\neg P\top). \tag{16}$$

Obviously, this statement can only be true in infinite models.

6 Soundness and Completeness of the Tableau Method

The following result states the adequacy of our tableau method.

Theorem 3 (Soundness and completeness). *Let $\varphi \in \mathcal{L}^s$. If φ is unsatisfiable, then the tableau for φ closes (completeness). If the tableau for φ closes then φ is unsatisfiable (soundness).*

6.1 Soundness

Soundness is proved using the notion of *interpretability.* A set of labeled formulae L is *interpretable* if there is a Kripke model $\mathcal{M} = (W, R, V)$ and a mapping $f : S \longrightarrow W$ such that for all $(\ell^E\ \varphi) \in L$, we have that $(W, R \setminus f(E), V), f(\ell) \models \varphi$, where $f(E) = \{(f(\ell_1), f(\ell_2)) \mid (\ell_1, \ell_2) \in E\}$. Then, we prove two facts:

Fact 2 *If φ is satisfiable, then, at any step of the construction of the tableau for φ, the set of labeled formulae of some branch is interpretable.*

Fact 3 *If φ is satisfiable, then any branch whose set of labeled formulae is interpretable cannot close. That is, there is an extension of this interpretable branch which does not close (this extended branch is therefore either open or infinite).*

These two facts combined together prove that, if φ is satisfiable, then the tableau for φ cannot close. That is, by contraposition, if the tableau for φ closes, then φ is unsatisfiable (soundness).

Proof (Fact 2). We prove the first fact by induction on the number of times n we use inference rules in the construction of the tableau for φ. The case $n = 0$ holds trivially: in that case L is a singleton $\{(\ell\ \varphi)\}$ and it suffices to define f so that it assigns to ℓ the world of the Kripke model where φ is satisfiable.

The induction step $n + 1$ is proved by examining each rule on a case by case basis and by showing that for each rule we can extend the mapping f associated to an interpretable branch to assign world(s) to the new label(s) created by the application of the rule and we can also extend the range of the accessibility relation R to assign a pair of worlds to the new relation term created by the rule. Indeed, we know by the Induction Hypothesis that there is a branch of the tableau whose terms are all interpretable by this mapping f.

The key steps to consider concern the rules $\Diamond$ and $\blacklozenge$.

- *Rule $\Diamond$:* Assume that the interpretable branch contains a labeled formula $(\ell^E\ \Diamond\varphi)$ (not chosen before in the execution of the tableau method). Applying the rule $\Diamond$ to this interpretable branch, we obtain $n + 1$ extended branches. We must show that one of them is interpretable. By assumption, we have that $(W, R \setminus f(E), V), f(\ell) \models \Diamond\varphi$. Therefore, there is $w \in W$ such that $f(\ell)Rw$ and $\mathcal{M}, w \models \varphi$. If w already corresponds to a label ℓ' such that $f(\ell') = w$, then we are in one of the first n extensions of the interpretable branch. In that case, the mapping f does not need to be extended and the label ℓ' has already been introduced in a rule earlier in the execution of the tableau method. Otherwise, we are in the last case of the rule $\Diamond$ and we need to extend the mapping f and assign to the 'fresh' label ℓ_{n+1} the possible world w: we set $f(\ell_{n+1}) := w$.

- *Rule $\blacklozenge$:* Assume that the interpretable branch contains a labeled formula $(\ell^E\ \blacklozenge\varphi)$ (not chosen before in the execution of the tableau method). Applying the rule $\blacklozenge$ to this interpretable branch, we obtain n extended branches corresponding to the n elements of $(M \times M) \cup \{(\ell_+, \ell_{++})\} \setminus E$.

We must show that one of them is interpretable. By assumption, we have that $(W, R \setminus f(E), V), f(\ell) \models \blacklozenge\varphi$. Therefore, there is a pair $(w, v) \in R \setminus f(E)$ such that $(W, R \setminus (f(E) \cup \{(w, v)\}), V), f(\ell) \models \varphi$. This pair of worlds (w, v) is either of the form $(f(\ell_0), f(\ell'_0))$, $(f(\ell_0), f(\ell_0))$, $(f(\ell_0), v)$, $(w, f(\ell_0))$, (w, w) or simply (w, v), for some labels ℓ_0, ℓ'_0 already introduced in this interpretable branch. The first five cases are covered by some of the cases corresponding to the elements of $M \times M$ of rule $\blacklozenge$ and the last case is covered by the case corresponding to (ℓ_+, ℓ_{++}) of rule $\blacklozenge$. So, at least one of the extended branches is interpretable and we can extend the mapping f accordingly.

Proof (Fact 3). We prove the second fact by contraposition. Assume that any extension of the initial interpretable branch closes. Then, any extension of this initial branch is such that it contains a labeled formula $(\ell^E\ \varphi)$ and its negation $(\ell^E\ \neg\varphi)$ or two labeled formulae $(\ell^E\ p)$ and $(\ell^F\ \neg p)$. This entails that the set of labeled formulae of any extended branch is *not* interpretable. So, by Fact 2, since φ is satisfiable by assumption, this entails that the set of labeled formulae of the *initial* branch is also not interpretable, which is impossible by assumption. □

Of course, much further combinatorial information is available in our tableaux than what has been employed so far, but we will not pursue this aspect here.

6.2 Completeness

We prove completeness by contraposition. Assume that the tableau for φ does not close. Then, there is an open branch in the tableau for φ. Let L be the set of labeled formulae appearing on this open branch and let T be the set of relation terms appearing on this open branch. We build the Kripke model $\mathcal{M} := (W, R, V)$ as follows:

- $W := \{\ell \mid (\ell^E\ \varphi) \in L$ for some $\varphi \in \mathcal{L}^s$ and $E \in 2^{S \times S}\}$;
- $R := \{(\ell_1, \ell_2) \mid (R\ \ell_1\ \ell_2) \in T\}$ and
- $V(p) := \{\ell \in W \mid (\ell^E\ p) \in L\}$, for all $p \in \mathbf{P}$.

Then, we have the following fact:

Fact 4 *For all labeled formulae $(\ell^E\ \chi)$,*

$$\textit{if } (\ell^E\ \chi) \in L \textit{ then } (W, R \setminus E, V), \ell \models \chi. \tag{17}$$

Proof. We prove Expression (17) by induction on the size of χ. The base case $\chi := p$ holds by definition of V. We prove the induction steps:

- $\chi := \varphi \wedge \psi$: Assume that $(\ell^E\ \varphi \wedge \psi) \in L$. Then, by saturation of the tableau rules, we also have that $(\ell^E\ \varphi)$ and $(\ell^E\ \psi)$ are in L. Then, by Induction Hypothesis, we must have that $(W, R \setminus E, V), \ell \models \varphi$ and $(W, R \setminus E, V), \ell \models \psi$. Hence, we obtain that $(W, R \setminus E, V), \ell \models \varphi \wedge \psi$.

- $\chi := \Diamond\varphi$: Assume that $(\ell^E \ \Diamond\varphi) \in L$. Then, by saturation of the tableau rules, there is $(R \ \ell \ \ell') \in T$ such that $(\ell, \ell') \notin E$ and $(\ell'^E \ \varphi) \in L$. Then, by Induction Hypothesis, $(W, R \setminus E, V), \ell' \models \varphi$ and $(\ell, \ell') \in R \setminus E$. Hence, we have that $(W, R \setminus E, V), \ell \models \Diamond\varphi$.
- $\chi := \blacklozenge\varphi$: Assume that $(\ell^E \ \blacklozenge\varphi) \in L$. Then, by saturation of the tableau rules, there is $(R \ \ell \ \ell') \in T$ such that $(\ell, \ell') \notin E$ and $(\ell^{E\cup\{(\ell,\ell')\}} \ \varphi) \in L$. Then, by Induction Hypothesis, $(W, R \setminus (E \cup \{(\ell, \ell')\}), V), \ell \models \varphi$ and $(\ell, \ell') \in R \setminus E$. Hence, we have that $(W, R \setminus E, V), \ell \models \blacklozenge\varphi$.
- $\chi := \neg p$: Assume that $(\ell^E \ \neg p) \in L$ and assume towards a contradiction that $(W, R \setminus E, V), \ell \models p$. Then, by definition of $\mathcal{M}$, there is a set of pairs of labels F such that $(\ell^F \ p) \in L$. However, if both $(\ell^E \ \neg p)$ and $(\ell^F \ p)$ belong to the same branch, the branch cannot be open, which is impossible by assumption.
- $\chi := \neg(\varphi \wedge \psi)$: Assume that $(\ell^E \ \neg(\varphi \wedge \psi)) \in L$. Then, by saturation of the tableau rules, we must have either that $(\ell^E \ \neg\varphi) \in L$ or $(\ell^E \ \neg\psi) \in L$. Then, by Induction Hypothesis, we must have either that $(W, R \setminus E, V), \ell \models \neg\varphi$ or $(W, R \setminus E, V), \ell \models \neg\psi$. In both cases, we obtain that $(W, R \setminus E, V), \ell \models \neg(\varphi \wedge \psi)$.
- $\chi := \neg\neg\varphi$: Assume that $(\ell^E \ \neg\neg\varphi) \in L$. Then, by saturation of the tableau rules, we obtain that $(\ell^E \ \varphi) \in L$. So, by Induction Hypothesis, we have that $(W, R \setminus E, V), \ell \models \varphi$, and therefore also $(W, R \setminus E, V), \ell \models \neg\neg\varphi$.
- $\chi := \neg\Diamond\varphi$: Assume that $(\ell^E \ \neg\Diamond\varphi) \in L$. Then, for all $(R \ \ell \ \ell') \in T$ such that $(\ell, \ell') \notin E$, $(\ell'^E \ \neg\varphi) \in L$ by saturation of the tableau rules. Therefore, by Induction Hypothesis, $(W, R \setminus E, V), \ell' \models \neg\varphi$, for all $(\ell, \ell') \in R \setminus E$. So, $(W, R \setminus E, V), \ell \models \neg\Diamond\varphi$.
- $\chi := \neg\blacklozenge\varphi$. Assume that $(\ell^E \ \neg\blacklozenge\varphi) \in L$. Then, by saturation of the rule $\neg\blacklozenge$ of the tableau rules, for all $(R \ \ell \ \ell') \in T$ such that $(\ell, \ell') \notin E$, we must have that $(\ell^{E\cup\{(\ell,\ell')\}} \ \neg\varphi) \in L$. So, by Induction Hypothesis, this entails that $(W, R \setminus (E \cup \{(\ell, \ell')\}), V), \ell \models \neg\varphi$. That is, $(W, R \setminus E, V), \ell \models \neg\blacklozenge\varphi$.

Thus, in particular, since $(\ell^\emptyset \ \varphi) \in L$ is the root of the tableau for φ, we have from Expression (17) that $\mathcal{M}, \ell \models \varphi$. Hence, φ is satisfiable.

This proves the completeness of our tableau method.

7 SML and Other Logics for Relation Change

In the remainder of this paper, we present no further concrete results about SML, but rather place sabotage modal logic against a broader background of logics for relation change, or more generally, model change. We hope that the points to be raised in this way lead toward a better understanding of a landscape of options here that is slowly emerging in the many systems available today.

What is special about SML? Sabotage modal logic has a number of features that make it like a modal logic. It is effectively axiomatizable via translation into first-order logic, even by means of a first-order tableau system, and its expressive power can be measured by invariance under a suitable notion of bisimulation. But perhaps surprisingly, given its simple-looking syntax and semantics, its complexity is high, validity being undecidable. What is the reason for this?

7.1 Dynamic-epistemic logic of relation change

One good way of approaching this issue is by comparison with another, and much more widely known modal logic of relation change, namely *dynamic-epistemic logic of relation transformers*. For a concrete example, think of the operation $|\varphi$ of 'link cutting' for a formula φ where, in the current accessibility relation R of a given model, we follow this instruction:

Remove all pairs (s, t) where s, t disagree on the truth value of φ.

The following result is well-known (cf. for instance [39]).

Fact 5 *The dynamic-epistemic logic of link cutting added on top of the basic modal logic is completely axiomatizable, and it is also decidable.*

The proof for this result goes by providing 'recursion axioms' in the typical DEL style, that show how to recursively push dynamic modalities inside through standard modalities. The key recursion axiom for link cutting looks as follows:

$$[|\varphi]\psi \leftrightarrow (\varphi \wedge \Box(\varphi \rightarrow [|\varphi]\psi) \vee (\neg\varphi \wedge \Box(\neg\varphi \rightarrow [|\varphi]\psi)))$$

Behind this result lies a much more general method, first presented in [39]. Any operation that replaces a relation R in a model by a new relation $\delta(R)$ (definable sub-relations are an important special case) automatically generates complete recursion axioms, provided that the transformation to $\delta(R)$ be definable in the format of a program in *propositional dynamic logic* PDL.[12]

So, what is the difference between the logics SML and DEL that accounts for this different complexity behavior?

Three contrasts. One clear difference is that DEL-style logics say precisely how the new relation is to arise form the old ones: it concerns *definable deletions*, as opposed to the arbitrary deletions of SML. Another contrast is that SML is about *global deletion* anywhere in the model, whereas one might think that things improve qua complexity when we work with *local relation change* only at the current distinguished point of the model. But there is yet one more distinction that may well turn out to be the most crucial one. Definable DEL-style relation change is *simultaneous*: all pairs not satisfying the given description are eliminated. By contrast, SML works *stepwise*: it removes links one by one.

In what follows we make a few observations on the last two contrasts, since the first one can be subsumed under these: SML performs stepwise deletions from the universal (and hence definable) relation $W \times W$.

7.2 Stepwise versions of DEL

Intuitively, the stepwise nature of SML is a source of complexity, since it blocks any obvious recursion axioms in the DEL style. The main reason is that dynamic

[12] For instance, for link cutting such a PDL program definition would look as follows: $(?\varphi; R; ?\varphi) \cup (?\neg\varphi; R; ?\neg\varphi)$.

modalities cannot be pushed through negations the way they are in DEL, since there can be many ways of performing an SML-deletion, whereas DEL style updates are usually (partial) *functions* allowing for modality/negation interchanges, as explained for instance in [36].

But also in more practical settings, it is well-known that changing classical simultaneous update scenarios such as the Puzzle of the Muddy Children into sequential ones (where children speak in turn) can completely change what happens in the long run, often even blocking any solution to the puzzle.

A good way of understanding the simultaneous/stepwise contrast is by importing the latter into the heartland of dynamic-epistemic logic. We could do this with the above link-cutting $|\varphi$, but an even simpler way of making the point concerns 'public announcement logic' PAL whose actions $!\varphi$ consists in removing all $\neg\varphi$-worlds from the current model, assuming that the formula φ is true in the distinguished 'actual world' of the current model. As above, PAL is completely axiomatizable using recursion axioms, and it is a decidable logic. Now let us introduce a new action $-\varphi$ into this system that works stepwise:

$\langle -\varphi\rangle\psi$ is true at $(\mathcal{M}, s)$ iff, after removing some $\neg\varphi$-point, ψ holds in s.

It is easy to see that this new system, let us call it PAL_{step}, changes the nature of PAL considerably. As explained above, there are no obvious recursion axioms, and even the basic modal invariance fails.

Fact 6 *PAL_{step} is not invariant for standard modal bisimulation.*

To see that this is so, just consider one model where point s has one successor t that is $\neg p$ and another model where s has two successors t, t' each having $\neg p$. There is an obvious modal bisimulation between the two models, but $\langle -\neg p\rangle\Diamond p$ is true in the second model, but not in the first.

Even so, PAL_{step} is clearly still translatable into first-order logic. However, its behavior in terms of validity is much more opaque than that of PAL itself. In fact, here is an obvious

Open Problem Is PAL_{step} decidable?

Truly new semantic methods may be needed for solving problems like this, and it may even be the case that PAL_{step} sides with SML rather than PAL qua complexity—and in that case, stepwiseness would override definability.

7.3 A local version of SML

Next, let us briefly consider another dimension of relation change, that might be thought to mitigate the high complexity. Instead of global versions, let us now look at local versions of our systems. For a start consider *loc*SML, a system that arises by specializing everything that we have defined for SML to deletions of links that start at the distinguished point of the current model.[13]

[13] A different type of 'local' sabotage is studied in [1,4], whose modality refers to the model transformation that deletes an outgoing edge from the evaluation point *and* moves the evaluation point to the target of the deleted arrow.

Here are a few observations to increase familiarity with this system.

Some validities All the principles listed earlier (Formulas (2) to (7)) as validities of SML are also valid in *loc*SML. Now we exhibit some differences.

Fact 7 *The following formula is valid in loc*SML*, but not in* SML*:*

$$\blacklozenge\Box\bot \to \blacksquare\Box\bot \tag{18}$$

And the following formula is valid in SML*, but not in loc*SML*:*

$$(\Diamond\Diamond\top \wedge \blacksquare\Box\bot) \to (\Box\varphi \to \varphi) \tag{19}$$

We leave the easy verifications to the reader. For instance, in the second formula, the global reading of the deletion modality $\blacksquare$ in the antecedent enforces that the current point is reflexive, whereas a local reading would not.

*Local loc*SML *vs. global* SML*.* The above remarks highlight the interest of the relation between global sabotage modal logic and its local variant. Within the compass of this paper, we merely provide a few observations.

In one direction, it seems that *loc*SML comes close to being inside SML. We have not been able to settle this in complete detail, but here is one observation about the special class of finitely-branching models.

Fact 8 *With finite branching, loc*SML *is invariant under sabotage bisimulation.*

Proof. Let Z be a sabotage bisimulation between two finitely-branching models $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$. The atomic and Boolean cases are routine. The Zig$_\Diamond$ and Zag$_\Diamond$ clauses of Definition 2 take care of $\Diamond$-formulae in a standard way familiar from basic modal logic. As for $\blacklozenge$-formulae (where $\blacklozenge$ denotes now the *local* variant of sabotage), assume that $\mathcal{M}, w \models \blacklozenge\varphi$. By the semantics of $\blacklozenge$, we have $\mathcal{M}, w \xrightarrow{\blacklozenge} \mathcal{M}_1, w \models \varphi$ where $\mathcal{M}_1$ is obtained via a *local* deletion and, by clause Zig$_\blacklozenge$ of Definition 2, it follows that $\mathcal{M}', w' \xrightarrow{\blacklozenge} \mathcal{M}'_1, w'$ and $(\mathcal{M}_1, w)Z(\mathcal{M}'_1, w')$. Here the second deletion could be non-local in general, but in our special class, it cannot be. There must be a local deletion that witnesses the transition to $\mathcal{M}'_1$, for otherwise the finite number of successors in $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$ would differ, which is not possible under the sabotage bisimulation that still obtains between $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$ – as we have shown in our discussion of the expressive power of SML. [14] By the induction hypothesis, we then conclude that $\mathcal{M}'_1, w' \models \varphi$ and, consequently, $\mathcal{M}', w' \models \blacklozenge\varphi$. Completely similarly, from $\mathcal{M}', w' \models \blacklozenge\varphi$, we conclude $\mathcal{M}, w \models \blacklozenge_l\varphi$ by clause Zag$_\blacklozenge$ of Definition 2. $\Box$

If we could show that *loc*SML is invariant for all sabotage bisimulations, then, by the obvious first-order translation of *loc*SML-formulas, our main characterization in Theorem 2 would imply that *loc*SML is embeddable into SML.

[14] This brute-force argument would fail in the presence of infinitely many successors.

Next, we consider the opposite direction. We show how to translate SML-formulas into *loc*SML-formulas, when some extra expressivity typical of hybrid logics [5] is added: the universal modality and the binder $\downarrow$. Let us first fix the translation (with the atomic case and Boolean clauses omitted):[15]

$$t(\blacklozenge\varphi) = \downarrow x.(x \wedge E\blacklozenge_l E(x \wedge t(\varphi)))$$

where E is the diamond of the universal modality, x is a state variable and $\blacklozenge_l$ denotes the local sabotage operator. It is easy to see that:

Fact 9 *For any formula φ of* SML*, model $\mathcal{M}$, and state w:*

$$\mathcal{M}, w \models \varphi \Longleftrightarrow \mathcal{M}, w \models t(\varphi).$$

where t is the translation defined above.

However, the price is high. It is known that standard modal logic plus the binder and the universal modality is equivalent to the first-order correspondence language itself (cf. [5]). Perhaps much weaker means would suffice. So, the direct relation from SML into *loc*SML remains unresolved.

Given this uncertainty about the precise relation between global and local sabotage logic, the complexity of reasoning in the latter also poses questions:

Open Problem Is *loc*SML decidable?

Our inclination is to doubt this, but we have not been able to give a proof.

7.4 Excursion: local DEL

Instead of pursuing the issues raised so far, we merely note that 'going local' need not be a force for simplicity in the other realm that we have contrasted with SML, namely, dynamic-epistemic logic DEL.

Suppose that we define a local version of link-cutting $|\varphi_{loc}$ as cutting links only between the current world and its accessible worlds. There are many concrete scenarios where this makes sense, for instance, when describing some local event of changing a communication link between agents in a network (cf. [33]).[16] Again we will see immediately that the typical DEL method of recursion axioms is going to fail. For, when we push a dynamic modality under a standard modality over successors of the current world, the link-cutting change is no longer local in these successors: it takes place somewhere else.

[15] From a hybrid logic point of view, we are now extending the hybrid logic known as $\mathcal{H}(E, \downarrow)$ with a local sabotage operator. Hybrid logics will return in Section 8.

[16] In [33] the authors assimilate this case to that of DEL-style deletions when the link to be deleted runs between two worlds having unique nominals as their names.

Localization Now, this problem can be solved, since, as in all logics that we are discussing here, all changes take place definably inside first-order logic. But in order to find recursion axioms for local link-cutting, or similar local relation-changing operations, we will have to be able to refer back to other worlds where local changes took place. In first-order terms, this means that we need a mechanisms for variable binding. In modal terms, we can use devices from *hybrid logics*, such as the universal modality, and especially, the 'binder', cf. $\downarrow$ [5]. It is a relatively standard exercise to provide recursion axioms for local link-cutting in a hybrid modal language, but we will not do so here. But there is no guarantee that such a hybrid extension of the modal base logic is still decidable, since hybrid logic with a universal modality plus downarrow is undecidable. [17]

In all, we conclude that introducing locality may be natural, but its complexity effects on logics of relation change are still ill-understood.

7.5 General logics of model change

We have identified a few general dimensions that affect design and complexity of logics for relation change. Clearly, there are many more systems of this sort than we could discuss here: even DEL itself has more sophisticated update mechanisms than those that we presented. But we hope to return to the more general structure of this landscape in follow-up work.

In this paper, we will just make a few references to further modal and first-order logics of relation change in Section 9, and give a first rough road map.

8 Sabotage Games and Fixpoint Logics

So far, we have looked at dynamic logics that describe single steps of model change, or more concretely, relation change. But the original sabotage game motivating a logic like SML was a many-step scenario unfolding over time, and these games have strategic global structure in the long run, going beyond local steps. What logic would naturally represent this structure? In this section we offer a few remarks that may clarify this issue, and show its interest.

8.1 The sabotage game

The sabotage game was introduced in [35], further studied in [31], and more recently in [22,40], and [37, Ch. 23]. It can be viewed as a game variant of a reachability problem where a player, Traveler, aims at reaching a predefined goal state (or goal region), while obstructed by the second player, Demon. We provide a short presentation of the game but we refer the reader to the above literature

[17] As we observed before, hybrid logic with a universal modality plus the binder $\downarrow$ is equivalent to the first-order correspondence language. It would be of interest to push things down to, say, hybrid logic with the 'at-operator' @ plus the binder, but even then we are in the undecidable 'Bounded Fragment' of the first-order language.

for more details and motivations, which range from studying the performance of algorithms under adverse circumstances to scenarios in learning theory.

The game is played on a (pointed-)frame (W, R) with a designated point w corresponding to the starting position of Traveler, and a designated non-empty subset $G \subseteq W$ (typically, a singleton for some designated point to be reached) corresponding to the 'goal' region of Traveler. Traveler moves locally by navigating the edges of R, one at the time, thereby constructing a path. Demon moves by deleting edges from R. The game proceeds in turns with Demon moving first. Traveler wins if she has reached her goal region. Otherwise Demon wins.

The sabotage game as defined above is a, possibly infinite, two-player, zero-sum, perfect information extensive form game. As such it is *determined*, that is, either Traveler or Demon has a winning strategy. In the case in which the graph W is finite, this determinacy is a consequence of Zermelo's theorem, [42]. In the general case, determinacy follows from the Gale-Stewart theorem, [21], as it is easy to see that the set of runs of the sabotage game where Traveler wins by reaching the goal is an open set in a standard topological sense. [18]

However, the case of reachability in standard directed graphs that has been paradigmatic for this paper trivializes the sabotage game. In the case in which Traveler's goal region is represented by a single state (that is, the original sabotage scenario), we have the following simple fact:

Fact 10 *Let (W, R) be a directed graph. If the goal region $G \subseteq W$ is a singleton not containing the initial position of Traveler, then the Demon has a winning strategy in the sabotage game played on (W, R).*

Proof. Unless Traveler started in the exit point already, Demon always has a winning strategy. That strategy works as follows. Demon cuts the link between the current position of the Traveler and the exit node if there is such a link, otherwise he cuts an arbitrary link. If Demon keeps doing this, Traveler never reaches the exist node. In a finite game, this is enough. On an infinite graph, this game can be infinite, but Demon will always produce infinite histories where Traveler does not reach the exit. This is enough. □

Obviously this trivialization does not go through if Traveler's goal region contains more than one state, perhaps defined by some goal formula, and this is the more general setting that we will assume in what follows. [19]

As we will see now, a natural extension of the language of SML yields the expressivity to neatly characterize winning positions in the sabotage game. It reflects a much more general point about game solution, developed at length in [37]: game-theoretic equilibria are definable in fixed-point logics for induction and recursion, and often even, in modal fixed-point logics.

[18] Cf. [37] on the background of these results in current logics of games.

[19] Other non-trivial variations arise when Traveler is allowed to move first, or if more than one link is allowed between points (the original sabotage game was played over 'multi-graphs'), or if we allow more than one graph relations, as in [31].

8.2 The sabotage μ-calculus: μSML

In the original paper [35], it was observed how the existence of Zermelo-style winning strategies can be expressed in a PDL-format. However, the natural more general format, covering also infinite sabotage games, is the following.

Syntax and semantics We extend the language of SML with a least fixpoint operator μ obtaining a sabotage version of the modal μ-calculus.

Syntax Let $\mathbf{P}$ be a countable set of propositional atoms. The set of formulae of the sabotage modal language $\mathcal{L}^{\mu}$ is defined by the following grammar in BNF:

$$\mathcal{L}^{\mu} : \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Diamond\varphi \mid \blacklozenge\varphi \mid \mu p.\varphi(p)$$

where $p \in \mathbf{P}$ and $\varphi(p)$ indicates that p occurs free in φ (i.e., it is not bounded by fixpoint operators) and under an even number of negations.[20] The remaining Boolean connectives and modal operators, as well as a greatest fix point operator, are defined in the standard way.

Semantics The satisfaction relation for $\mathcal{L}^{\mu}$ is defined as for $\mathcal{L}^{s}$. The semantics of the fixpoint operator is also standard (subject to the caveats in Footnote 18):

$$\mathcal{M}, w \models \mu p.\varphi(p) \Longleftrightarrow w \in \bigcap \left\{ X \in 2^{W} \mid ||\varphi||_{\mathcal{M}[p:=X]} \subseteq X \right\}$$

where $||\varphi||_{\mathcal{M}[p:=X]}$ denotes the truth-set of φ once $V(p)$ is set to be X.

Our earlier standard translation from SML into FOL (Definition 1) can be extended in a natural way to yield a translation of μSML formulas into FO(LFP), the well-known extension of FOL with least fixpoint operators.

As the only higher-order quantification involved in μSML is over monadic predicates (propositions), we can also obtain a translation into monadic second-order logic (MSO). Which fragment of MSO (or of FO(LFP)) is precisely identified by μSML through the standard translation is an open question. However, we can establish the following simple necessary condition:

[20] This syntactic restriction guarantees that every formula $\varphi(p)$ defines a monotonic set transformation that preserves $\subseteq$, which in turn guarantees the existence of least and greatest fixpoints by the Knaster-Tarski theorem. We refer the reader to [41] for an excellent introduction to the standard modal μ-calculus. While this sounds totally routine, it should be said that the monotonicity now needs an additional check, namely, that the new sabotage modalities, which refer to changed models, preserve it. The semantic argument for this is straightforward, but we do note that this need not always be the case. For instance, in a μ-calculus with explicit dynamic-epistemic modalities such as $[!\varphi]\psi$, positive occurrences of p in ψ present no problem, but such occurrences in φ do not guarantee monotonicity, and one needs to restrict the syntax of ψ further, say, to existential modal formulas preserved under extending models.

Fact 11 *μSML formulas are invariant under sabotage bisimulation.*

Proof. The fact follows directly from the proof of Proposition 2 plus the fact that formulas involving the least fixpoint operator μ can be 'unpacked' in a standard manner into infinitary disjunctions. □

8.3 Expressing winning regions in the sabotage game

Now let us descend from the general to the particular. Our earlier observations prose simple tests for putative logics of sabotage games. In particular, Fact 10 suggests that the original sabotage modal logic SML, possibly with some simple extra resources, should be able to express the existence of Demon's winning strategy as a validity. Here is such a formula, using two additional modal devices:

$$U((\neg\texttt{goal} \wedge \Diamond\top) \rightarrow \blacklozenge\Box\neg\texttt{goal}) \tag{20}$$

where U is the universal modality and $\texttt{goal}$ is a nominal for the goal region.

But it is μSML which has the resources to express the general sets of winning positions for Traveler (and Demon) as formulas, even with non-singleton winning regions, provided these are defined by formulas $\texttt{goal}$ of our language.

Fact 12 *The set of winning positions for Traveler in the sabotage game on (W, R) is defined by the following formula of μSML:*

$$\texttt{win} := \mu p.\texttt{goal} \vee (\Diamond\top \wedge \blacksquare\Diamond p) \tag{21}$$

where $\texttt{goal}$ defines the goal region G.

Proof. Let $\mathcal{M}$ be a model defined on (W, R). By the semantics of μ the claim amounts to the following equation:

$$||\texttt{win}|| = \bigcap \left\{ X \in 2^W \mid ||\texttt{goal} \vee (\Diamond\top \wedge \blacksquare\Diamond p)||_{\mathcal{M}[p:=X]} \subseteq X \right\}$$

This denotes the smallest set of states X such that:

$$X = ||\texttt{goal} \vee (\Diamond\top \wedge \blacksquare\Diamond p)||_{\mathcal{M}[p:=X]}$$

That is the smallest X corresponding to the set of states in which either Traveler has reached her goal states or she can respond to all deletions by Demon by accessing X itself. It follows that for each in X there exists a strategy for Traveler that enables her to reach one of her goal states, that is, a winning strategy. □

Incidentally, by the determinacy of the sabotage game, the dual formula of (21) defines the winning region for Demon, which is therefore the largest fixpoint of the equation $p \leftrightarrow \neg\texttt{goal} \wedge (\Box\bot \vee \blacklozenge\Box p)$. Viewed in this light, the earlier Formula (20) for Demon stated that $\neg\texttt{goal}$ is such greatest fixpoint.

9 Related Work

As we have noted right at the start, SML has been studied extensively in a recent series of works [1,3,2,4] and [17], parallel to ours but with wider scope. In particular [17,4] established that the model-checking problem of SML is PSPACE-complete, that SML lacks the finite model property, the tree-model property and that its satisfiability problem is undecidable. This considerably improved the same results obtained for the multi-modal variant of SML in [31,30].

The cited authors also introduce a tableau method for SML in [2]. They use a special system of nominals in their tableaux, which represent possible worlds. This allows more control over the application of the tableau rules than the calculus in this paper. Indeed, in the rules $\blacklozenge$ and $\lozenge$, instead of considering all possible combinations of nominals as we do, their method uses a refinement of the usual diamond rule of modal logic with nominals, together with an extra rule called (ub) that suitably 'controls' the combination of all the nominals. Even though these features may make the tableaux in [2] slightly less transparent in terms of how the rules relate to the underlying semantics of SML, it may well make the system more efficient than ours from a computational standpoint.

Often triggered directly by the initial work on dynamic-epistemic logic and sabotage modal logic, many further logics of graph modifiers have been proposed in the recent literature. Relevant systems include local and global graph modifiers [7], dynamic epistemic modifiers [11,10], dynamic modal logic DML [15], arrow logic [28], logics of copy and remove [6], the logic of preference upgrade, [39] and general dynamic dynamic logic [23].[21]

Instead of discussing this terrain in detail, we give a small 'expressiveness map' based on what can readily be observed, and on what authors have claimed.[22] In Figure 7, an arrow $L_1 \to L_2$ means that logic L_1 is strictly more expressive than L_2. When there is no arrow between two logics, they are incomparable in terms of expressiveness. Here is how one can read the map.

At the bottom level, there is just basic modal logic, which is expressively equivalent to dynamic modal logic DML (cf. [15]), arrow logic (cf. [28]) and preference upgrade logic (cf. [39]). One step up, at the middle level of our map, we find SML and the other logics inspired by it later, such as swap logic, bridging logic, and local sabotage logic in the sense of [1,3]. [23] From the cited publications, we know that these mid-level systems are incomparable in terms of expressivity, and that they are also strictly more expressive than the basic modal logic (cf. [4])

[21] Also relevant here is the work in Gabbay [20] on 'reactive logics', where the accessibility relation of modal models is changed during the interpretation process of formulas. The latter procedure is reminiscent of 'dynamic semantics' in the Amsterdam style, but we leave a precise comparison for future work.

[22] Here we disregard logics with Kleene $*$ or fixed-point extensions, where things can be different. In particular, we do not treat general dynamic dynamic logic [23], which is equivalent to PDL. Also, we were unable to fit in the logics of copy and remove.

[23] As observed earlier (footnote 13) the system of 'local' sabotage from [1,3] is subtly different from the one we considered in this paper.

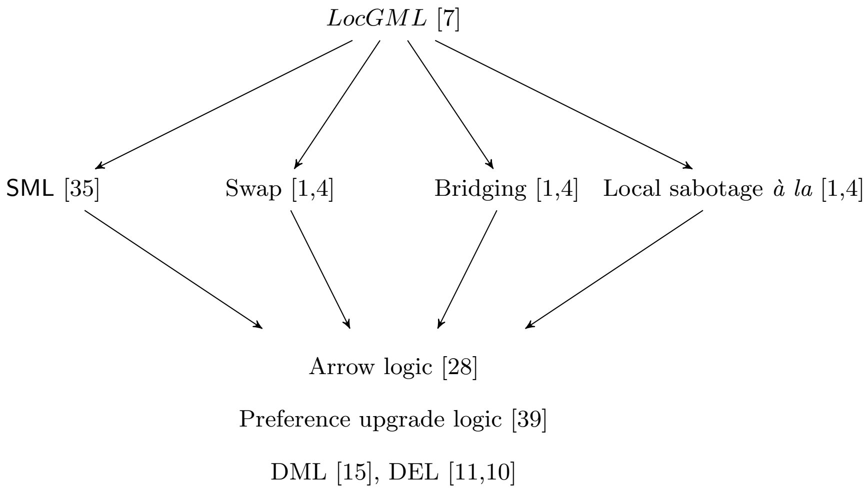

Fig. 7. Expressiveness map of some dynamic modal logics for graph modifiers.

and at the same time, strictly less expressive than the full hybrid language with universal modality plus binders (cf. [2]). Incidentally, some of these intermediate systems (e.g., swap logic) have been also shown to be strictly less expressive than the so-called Bounded Fragment of FOL, which is undecidable, but still natural, cf. [3]. Finally, the logic of local and global graph modifiers proposed in [7] arose from combining features of all these systems, and it has been shown to be at least as expressive as hybrid logic with the universal modality plus binders, that is, the full FOL correspondence language.

Of course, comparing logics by expressiveness like this is a drastic projection of their different dynamic motivations, and our map is only a first step. We can compare these systems in other ways, and in particular, the stated equivalence are all up for grabs if we generalize to temporal protocol models, where the usual reduction axioms become invalid. We leave a more sensitive and detailed analysis of the area of logic for model change to future work.[24]

10 Conclusions and Future Work

In this paper, we have dusted off sabotage modal logic SML, and looked at its broader current relevance. We provided a first-order translation for this system,

[24] Our discussion here has been model-theoretic, in terms of semantic invariances and expressive power. But there are also other ways of achieving generality in the rich and growing landscape of dynamic logics for model change. Alternatively, one could use algebraic methods, category-theoretic methods, or proof-theoretic perspectives, which might suggest systematizations of their own. One such alternative approach that we would like to mention specifically for its general power, is the framework of 'update logic' [9], based on display-style substructural proof theory.

proved a novel characterization theorem in terms of a new notion of sabotage bisimulation, and introduced a sound and complete tableau system for validity.

Still, SML remains an under-investigated system and many natural questions remain open. For instance, we noted that the schematic validities of SML form a natural subset whose axiomatizability, or decidability, is an open problem. Another natural theme would be modal correspondence theory, whose techniques are known to extend into higher-order logic [34]. In particular, can the Sahlqvist theorem be generalized to SML? Also, in our investigation, we arrived at natural variations of SML, such as its version with only local deletions, whose decidability is still open.[25] Also, given the apparent complexity of the system, it would make sense to look at natural fragments, such as the closed formulas only, that formed many of our examples, or the formulas of sabotage depth 1.

Next, we have presented sabotage modal logic as a member of a species, that of modal logics describing model changes. We have identified some of its basic features, and in particular, its arbitrary non-definable and stepwise character, setting it apart, for instance, from the better-studied world of dynamic-epistemic logics. We see our way of looking as the start of something more ambitious, finding a better map of the whole field of logics for model change, and the major parameters that determine their complexity. Here, too, we encountered many specific open problems, as the complexity of dynamic-epistemic logics extended with stepwise update modalities in the SML-style. But to us, the greatest challenge would be a general perspective on dynamic logics of model change that leads to general theorems beyond those for specific logics. We made some suggestions to this effect in Section 8, but clearly this was just a start.

Finally, returning to the sabotage games that motivated SML in the first place, we introduced sabotage μ-calculus as a system for defining strategic powers of players, but also as a testing ground for what happens when we introduce model-changing modalities that can affect the process of approximation leading to the usual smallest or greatest fixed-points. The sabotage μ-calculus system sat inside first-order fixed-point logic FO(LFP) and monadic second-order logic MSO, but it seems a very natural object of study in its own right.

But perhaps more importantly than just studying logics of games, sabotage games themselves remain under-explored as providing the real dynamic scenarios unfolding over time, beyond the logics of single-step model change discussed above. There are many natural variations on the original scenario in [35][26]— and these may well come to serve as pilots for the newly emerging logical study of games played over networks, such as the argumentation games of [24,19], or

[25] Other natural variations include multi-modal SML with indices for the relations to be cut from [31,30], or logics of adding links rather than deleting them, like the aforementioned modal logics of bridging [4]. In this connection, one may note that adding links to R is the same as deleting links from the complement $-R$, so there are connection laws to be had.

[26] One alternative scenario already suggested in [35] is that of a three-player game where Demon tries to prevent two Travelers from meeting. This involves moves in a product graph that do not fall directly under our earlier analysis.

the social network games suggested in [38] and [14]. Eventually, of course, such network games need not be about 'sabotage' at all, and the general area that we are in becomes that of games that change their own playground, for which dynamic logics of model change provide a backdrop, but not the whole story.

Acknowledgments. For this research Davide Grossi was supported in part by NWO (VENI grant 639.021.816), and in part by EPSRC (grant EP/M015815/1). Guillaume Aucher was supported in part by AFR (grant TR-PDR BFR08-056). Johan van Benthem was supported in part by the Changjiang Scholars program of the Chinese Ministry of Education. The authors would like to thank Valentin Goranko for insightful discussions on the topic of this paper.

References

1. C. Areces, R. Fervari, and G. Hoffmann. Moving arrows and four model checking results. In L. Ong and R. Queiroz, editors, *Logic, Language, Information and Computation*, volume 7456, pages 145–153, 2012.
2. C. Areces, R. Fervari, and G. Hoffmann. Tableaux for relation-changing modal logics. In P. Fontaine, C. Ringeissen, and R. Schmidt, editors, *Proceedings of the 9th International Symposium on Frontiers of Combining Systems (FroCoS'13)*, volume 8152 of *LNAI*, pages 263–278, 2013.
3. C. Areces, R. Fervari, and G. Hoffmann. Swap logic. *Logic Journal of the IGPL*, 22(2):309–332, 2014.
4. C. Areces, R. Fervari, and G. Hoffmann. Relation-changing modal operators. *Logic Journal of the IGPL*, 2015.
5. C. Areces and B. Ten Cate. Hybrid logics. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *Handbook of Modal Logic*, pages 821–868. Elsevier, 2006.
6. C. Areces, H. van Ditmarsch, R. Fervari, and F. Schwarzentruber. Logics with copy and remove. In *Logic, Language, Information and Computation*, volume 8652, pages 51–65, 2014.
7. G. Aucher, P. Balbiani, L. Fariñas del Cerro, and A. Herzig. Global and local graph modifiers. *Electronic Notes in Theoretical Computer Science*, 231:293–307, 2009.
8. G. Aucher, J. van Benthem, and D. Grossi. Sabotage modal logic: Some model and proof theoretic aspects. In *Proceedings of LORI 2015*, 2015.
9. Guillaume Aucher. Displaying updates in logic. *Journal of Logic and Computation*, 2016.
10. Alexandru Baltag and Lawrence S. Moss. Logics for epistemic programs. *Synthese*, 139(2):165–224, 2004.
11. Alexandru Baltag, Lawrence S. Moss, and Slawomir Solecki. The logic of public announcements and common knowledge and private suspicions. In Itzhak Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-98), Evanston, IL, USA, July 22-24, 1998*, pages 43–56. Morgan Kaufmann, 1998.
12. P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic.* Cambridge University Press, Cambridge, 2001.
13. C. C. Chang and H. J. Keisler. *Model Theory.* Studies in Logic and the Foundations of Mathematics. North-Holland, 1973.

14. Z. Christoff. *Dynamic Logics of Networks*. PhD thesis, ILLC, University of Amsterdam, 2016.
15. Gerard R. Renardel de Lavalette. Changing modalities. *J. Log. Comput.*, 14(2):251–275, 2004.
16. M. de Rijke. A note on graded modal logic. *Studia Logica*, 64(2):271–283, 2000.
17. R. Fervari. *Relation-Changing Modal Operators*. PhD thesis, Universidad Nacional de Córdoba, Facultad de Matemática, Astronomía y Física, 2013.
18. M. Fitting. *Fitting, M.* Reidel, 1983.
19. D. Gabbay and D. Grossi. When are two arguments the same? equivalence in abstract argumentation. In A. Baltag and S. Smets, editors, *Johan van Benthem on Logic and Information Dynamics*. Springer, 2014.
20. Dov M. Gabbay. *Reactive Kripke Semantics*. Cognitive Technologies. Springer, 2013.
21. D. Gale and F. M. Stewart. Infinite games with perfect information. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games II*, volume 28 of *Annals of Mathematics Studies*, pages 245–266. Princeton University Press, 1953.
22. N. Gierasimczuk, L. Kurzen, and F. Velázquez-Quesada. Learning and teaching as a game: a sabotage approach. In X. He, J. Horty, and E. Pacuit, editors, *Proceedings of LORI-2*, number 5834 in LNAI, 2009.
23. Patrick Girard, Jeremy Seligman, and Fenrong Liu. General dynamic dynamic logic. In Thomas Bolander, Torben Braüner, Silvio Ghilardi, and Lawrence S. Moss, editors, *Advances in Modal Logic 9, papers from the ninth conference on "Advances in Modal Logic," held in Copenhagen, Denmark, 22-25 August 2012*, pages 239–260. College Publications, 2012.
24. D. Grossi. On the logic of argumentation theory. In W. van der Hoek, G. Kaminka, Y. Lespérance, and S. Sen, editors, *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, pages 409–416. IFAAMAS, 2010.
25. D. Grossi. Argumentation theory in the view of modal logic. In P. McBurney and I. Rahwan, editors, *Post-proceedings of the 7th International Workshop on Argumentation in Multi-Agent Systems*, number 6614 in LNAI, pages 190–208, 2011.
26. S. Gruener, F. Radmacher, and W. Thomas. Connectivity games over dynamic networks. *Theoretical Computer Science*, 498:46–65, 2013.
27. W. Holliday, T. Icard, and T. Hoshi. Schematic validity in dynamic epistemic logic: Decidability. In H. van Ditmarsch, J. Lang, and S. Ju, editors, *Proceedings of the Third International Workshop on Logic, Rationality and Interaction (LORI'11)*, volume 6953 of *LNAI*, pages 87–96, 2011.
28. B. Kooi and B. Renne. Arrow update logic. *Review of Symbolic Logic*, 4(4), 2011.
29. F. Liu, J. Seligman, and P. Girard. Logical dynamics of belief change in the community. *Synthese*, 191(11):2403–2431, 2014.
30. C. Löding and P. Rohde. Model checking and satisfiability for sabotage modal logic. In P. K. Pandya and J. Radhakrishnan, editors, *FSTTCS 2003*, volume 2914 of *LNCS*, pages 302–313. Springer, 2003.
31. C. Löding and P. Rohde. Solving the sabotage game is PSPACE-hard. Technical report, Department of Computer Science RWTH Aachen, 2003.
32. F. Radmacher and W. Thomas. A game theoretic approach to the analysis of dynamic networks. *Electronic Notes in Theoretical Computer Science*, 200(2):21–37, 2008.

33. J. Seligman, F. Liu, and P. Girard. Facebook and epistemic logic of friendship. In *Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge (TARK'13)*, pages 229–238, 2013.
34. J. van Benthem. *Modal Logic and Classical Logic*. Monographs in Philosophical Logic and Formal Linguistics. Bibliopolis, 1983.
35. J. van Benthem. An essay on sabotage and obstruction. In D. Hutter and W. Stephan, editors, *Mechanizing Mathematical Reasoning*, volume 2605 of *LNCS*, pages 268–276. Springer, 2005.
36. J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2011.
37. J. van Benthem. *Logic in Games*. MIT Press, 2013.
38. J. van Benthem. Oscillation, logic and dynamical systems. In J. Szymanik and S. Gosh, editors, *The Facts Matter. Essays on Logic and Cognition in Honour of Rineke Verbrugge*. College Publications, 2015.
39. J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, 17(2), 2007.
40. F. Velázquez-Quesada. *Small Steps in the Dynamics of Information*. PhD thesis, ILLC, University of Amsterdam, 2011.
41. Y. Venema. Lectures on the modal μ-calculus. Renmin University in Beijing (China), 2008.
42. E. Zermelo. Über eine anwendung der mengenlehre auf die theorie des schachspiels. In *Proceedings of the 5th Congress Mathematicians*, pages 501–504. Cambridge University Press, 1913.