Modal logics of sabotage revisited

Guillaume Aucher, Johan Van Benthem, Davide Grossi

▶ **To cite this version:**

Guillaume Aucher, Johan Van Benthem, Davide Grossi. Modal logics of sabotage revisited. Journal of Logic and Computation, 2018, 28 (2), pp.269 - 303. 10.1093/logcom/exx034 . hal-01827076

HAL Id: hal-01827076
https://inria.hal.science/hal-01827076

Submitted on 1 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modal Logics of Sabotage Revisited

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

[1] Univ Rennes, CNRS, IRISA
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 graphs they are played on. We investigate some model-theoretic and proof-theoretic aspects of sabotage modal logic, which has come to be viewed as an early dynamic logic of graph change. Our first result 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'. Next, we offer a sound and complete tableau method and its associated labeled sequent calculus 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 introduced in [44], expands basic modal logic with a modality $\blacklozenge\varphi$ saying "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 start of a line of systems in a dynamic-epistemic spirit [46], such as 'graph modifier logic' [8], 'swap logic' [3], or 'arrow update logic' [31,1]. SML is also directly related to recent work in theoretical computer science [39,29], learning theory [24], logics of social networks [42,32], and argumentation [26,27].

Only a few properties of sabotage modal logic have been studied in depth so far. The source paper [44] showed how SML can formulate solution concepts for sabotage games, it gave a SPACE upper bound for model checking, and it sketched how the SML language can be translated into first-order logic, though the validities are not closed under substitutions. More incisive results are in [34,33], where multi-modal SML is shown to have an undecidable satisfiability problem and a PSPACE-complete model-checking problem. A main open problem mentioned in [34,33] is a bisimulation invariance characterization for SML. This was solved independently in the proceedings version of the present article [9] and in the work of an Argentinean team [4]. Also independently, both teams developed tableau systems to analyze the structure of validity in SML, cf. [2,4]. We will discuss some connections in Section 8.1 at the end of this paper.

Outline Our paper has two parts. The first, comprising Sections 2 to 5, extends the material in [9] with proofs left out there, while adding a labeled sequent calculus, several new examples and a few small novel observations. The two main contributions are 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 and matching sequent calculus in Section 5. The second part of the paper is more exploratory. Sections 6 to 7 present new material addressing the issue of what makes SML tick among current modal logics of model change, where we identify 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 fixed-point logics of sabotage for analyzing graph games, and revisit the original motivation for SML in sabotage games. Section 8 discusses related work. Section 9 concludes and emphasizes once more how sabotage modal logics can serve as a laboratory for broader graph games.

2 Preliminaries

In this section, we introduce the syntax and semantics of SML, recapitulate some key results from [33], and present a standard translation from SML into FOL.

2.1 Syntax and semantics

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

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

where $p \in \mathbf{P}$. We add some useful notation. For all $\varphi, \psi \in \mathcal{L}^s$, we define the usual abbreviations: $\blacksquare\varphi \triangleq \neg\blacklozenge\neg\varphi$, $\varphi \vee \psi \triangleq \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi \triangleq \neg\varphi \vee \psi$, $\bot \triangleq p \wedge \neg p$, $\top \triangleq \neg\bot$ (for an arbitrary $p \in \mathbf{P}$). For all $\varphi \in \mathcal{L}^s$, we define iterated formulas as follows: for all $\mathrm{M} \in \{\blacklozenge, \Diamond, \blacksquare, \Box\}$, $\mathrm{M}_0\varphi \triangleq \varphi$ and for all $n \in \mathbb{N}$, $\mathrm{M}_{n+1}\varphi \triangleq \mathrm{M}\mathrm{M}_n\varphi$. To save parentheses, binding strength is in this order: $\blacksquare, \blacklozenge, \Box, \Diamond, \neg, \wedge, \vee, \rightarrow$. Also, we will use the 'sabotage depth' $sd(\varphi)$ of formulas φ, [33], defined inductively by: $sd(p) \triangleq 0$, $sd(\neg\varphi) \triangleq sd(\varphi)$, $sd(\varphi_1 \wedge \varphi_2) \triangleq \mathtt{max}\{sd(\varphi_1), sd(\varphi_2)\}$, $sd(\Diamond\varphi) \triangleq sd(\varphi)$ and $sd(\blacklozenge\varphi) \triangleq sd(\varphi) + 1$. Hence, in particular, $sd(\top) = 0$.

We work with standard relational models $\mathcal{M} = (W, R, V)$ for modal logic ('models', for short), with W a non-empty set; $R \subseteq W \times W$, and $V : \mathbf{P} \longrightarrow 2^W$. A pair $\mathcal{F} = (W, R)$ is called a frame. We write $w \in \mathcal{M}$ for $w \in W$, and call a pair $(\mathcal{M}, w)$ a 'pointed model', the class of which is $\mathfrak{M}$. The satisfaction relation $\models \subseteq \mathfrak{M} \times \mathcal{L}^s$ is defined inductively by truth conditions. Those for the Boolean and modal connectives $\neg, \wedge, \Diamond$ are as usual, and the truth condition for $\blacklozenge$ is:

$$(W, R, V), w \models \blacklozenge\varphi \text{ iff there is } (u, v) \in R \text{ s.t. } (W, R \setminus \{(u, v)\}, V), w \models \varphi.$$

Thus, $\blacklozenge\varphi$ is true in a pointed model if there are two R-related (possibly identical) states such that, if the edge between these states is removed from R, φ holds

at the distinguished state. (If no such pair exists, $\blacklozenge\varphi$ is false for any φ.) The triple $\mathsf{SML} \triangleq (\mathcal{L}^s, \mathfrak{M}, \models)$ is called *sabotage modal logic*. Satisfiability, validity and logical consequence for SML are defined as usual.

We can also restate this slightly differently to emphasize the bimodal flavor of our logic. Define a relation $\mathbf{r} \subseteq \mathfrak{M} \times \mathfrak{M}$ with $((W, R, V, w), (W', R', V', w')) \in \mathbf{r}$ iff $W' = W$; $R' = R \setminus \{(u, v)\}$ for some $u, v \in W$; $V' = V$ and $w' = w$. Then the above truth condition for the sabotage modality makes it a standard existential modality referring to this ordering between models. We can also iterate this order in the obvious way, to talk about models reachable in finitely many $\mathbf{r}$-steps, obtaining the relation $\mathbf{r}^*$.

As usual, our standard relational models can also be interpreted as models for the binary fragment of FOL with equality[1] denoted $\mathcal{L}^1$. Sometimes in the proofs to follow we will use the following first-order terminology and 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 – or 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 (also called a 'type') if there exists an element $w \in W$ such that $\mathcal{M} \models \Gamma(x)[w]$.

If $(\mathcal{M}, w) \in \mathfrak{M}$, the *sabotage modal theory* of $(\mathcal{M}, w)$ is the set $\mathbb{T}^s(\mathcal{M}, w) \triangleq \{\varphi \in \mathcal{L}^s \mid \mathcal{M}, w \models \varphi\}$. We also define the binary relation $\leftrightsquigarrow_s \subseteq \mathfrak{M} \times \mathfrak{M}$ as follows: $(\mathcal{M}, w) \leftrightsquigarrow_s (\mathcal{M}', w')$ (for $((\mathcal{M}, w), (\mathcal{M}', w')) \in \leftrightsquigarrow_s$) iff for all $\varphi \in \mathcal{L}^s$, $\mathcal{M}, w \models \varphi$ iff $\mathcal{M}', w' \models \varphi$. In that case, we say that $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$ are *sabotage modally equivalent* (that is, they satisfy the same sabotage modal formulas).

2.2 Some notable validities and expressible properties

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

$$\blacksquare\bot \to \Box\bot \tag{1}$$

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

$$p \to \blacksquare p \tag{3}$$

$$(\blacksquare p \wedge \neg\blacksquare\bot) \to p \tag{4}$$

$$\Box p \to \blacksquare(\Diamond\top \to \Diamond p) \tag{5}$$

$$\Diamond\varphi \wedge \Diamond\neg\varphi \to \blacklozenge\blacklozenge\top \tag{6}$$

The fact that we are using propositional atoms instead of variables for formulas in the first five of the above validities 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 (5). Consider its schematic formulation

$$\Box\varphi \to \blacksquare(\Diamond\top \to \Diamond\varphi), \tag{7}$$

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

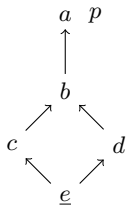

Fig. 1. A model showing that the general schema $\Box\varphi \to \blacksquare(\Diamond\top \to \Diamond\varphi)$ fails in SML. Let $\varphi \triangleq \Diamond\Diamond p$, let $V(p) = \{a\}$, and let the evaluation point be e.

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 (6) is a schematic validity.[2] Still, any logic has a 'substitution core' consisting of its schematically valid principles, cf. [46], and it is an interesting open problem whether SML has a decidable, or even an axiomatizable substitution core.[3]

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 $1 \leq n$ (Expression (8), but see also Example 2 of Section 4 below); "there exist at least two successors or, there exists at most one successor and at least another state with at least one successor" (Expression (9)); "there exists no successor but there is at least one state with at least one successor" (Expression (10)). All these properties have FOL formulations using counting quantifiers:

$$\exists^{\leq n} y\,(xRy) \tag{8}$$

$$\exists^{\geq 2} y\,(xRy) \vee \big(\exists^{\leq 1} y\,(xRy) \wedge \exists y(x \neq y \wedge \exists z(yRz)\big) \tag{9}$$

$$\neg\exists y(xRy) \wedge \exists y(\exists z(yRz)) \tag{10}$$

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

$$\Box\bot \vee \bigvee_{1\leq i\leq n} \blacklozenge_i\Box\bot \tag{11}$$

$$\blacklozenge\Diamond\top \tag{12}$$

$$\Box\bot \wedge \blacklozenge\top \tag{13}$$

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].

[2] A formula φ of a language $\mathcal{L}$ is a schematic validity on a class of models if $\sigma(\varphi)$ is valid in that class for any substitution σ of arbitrary formulas for proposition letters.

[3] The analogous open problem for the much simpler dynamic-epistemic logic PAL of public announcement was only solved in [30]: it turned out to be decidable.

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

$$\Diamond\top \wedge \Box\Diamond\top \wedge \blacksquare\Box\bot \tag{14}$$

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:

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

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:

$$CYCLE(1) \triangleq PATH(1) = (14)$$
$$CYCLE(n+1) \triangleq PATH(n) \wedge \bigwedge_{1\leq i\leq n} (\neg\Diamond_{n-i}\blacklozenge_{n-i} CYCLE(i))$$

The base case is immediate. As to the induction step, the first conjunct forced existence of a unique path of length at least n from the current state. The second conjunct forces the $(n-1)^{th}$-step in such a path to end in the current state. □

2.3 A First-Order Translation for SML

A standard first-order translation for SML was sketched in [44] and [34]. Later on, a detailed translation was independently given in the proceedings version of this paper [9] 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 is achieved by indexing the standard translation with a set E of pairs of variables. When the translation is applied to the outermost operator of a given formula, this set is empty. As analysis proceeds towards inner operators, each sabotage operator $\blacklozenge$ in the formula will introduce a new pair of variables in E, which is 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:*

$$ST_x^E(p) \triangleq P(x) \qquad ST_x^E(\bot) \triangleq \neg(x=x)$$

$$ST_x^E(\neg\varphi) \triangleq \neg ST_x^E(\varphi) \qquad ST_x^E(\varphi_1 \wedge \varphi_2) \triangleq ST_x^E(\varphi_1) \wedge ST_x^E(\varphi_2)$$

$$ST_x^E(\Diamond\varphi) \triangleq \exists y \left(xRy \wedge \bigwedge_{(v,w)\in E} \neg(x=v \wedge y=w) \wedge ST_y^E(\varphi) \right)$$

$$ST_x^E(\blacklozenge\varphi) \triangleq \exists y,z \left(yRz \wedge \bigwedge_{(v,w)\in E} \neg(y=v \wedge z=w) \wedge ST_x^{E\cup\{(y,z)\}}(\varphi) \right)$$

In the inductive clauses, a standard modality $\Diamond\varphi$ translates as a first-order formula with x free, stating there exists a state y accessible from x via an edge different from all edges in the set E, with the translation of φ holding at y. Formula $\blacklozenge\varphi$ becomes a first-order formula saying there is some R-edge (y,z) different from any edge already in E, that the translation of φ should now be taken with respect to the set $E\cup\{(y,z)\}$, and that the result holds at x. Setting up the translation like this lets one book-keep removal of edges in a perspicuous manner via E, forcing modalities to access ever smaller relations.

It is important to notice the following feature of this procedure. Depending on the chosen E, ST^E can possibly yield formulas with several free variables, e.g.: $ST_x^{\{(v,w)\}}(\Diamond p) = \exists y\,(xRy \wedge \neg(x=v \wedge y=w) \wedge P(y))$. 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 subformulas 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:

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

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* $\varphi \in \mathcal{L}^s$*:*

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

Proof. This goes by induction on the structure of φ. The key inductive step for the sabotage operator $\blacklozenge$ is proven by the equivalences that use the inductive hypothesis plus the semantics of $\blacklozenge$ and the definition of our translation ST:

$$\begin{aligned}
\mathcal{M}, w \models \blacklozenge\varphi \text{ iff } & \text{ there exists } \mathcal{M}' \text{ with } (\mathcal{M}, w)\mathbf{r}(\mathcal{M}', w) \text{ and } \mathcal{M}', w \models \varphi \\
\text{iff } & \text{ there exists } \mathcal{M}' \text{ with } (\mathcal{M}, w)\mathbf{r}(\mathcal{M}', w) \text{ and } \mathcal{M}' \models ST_x^{\emptyset}(\varphi)[w] \\
\text{iff } & \mathcal{M} \models ST_x^{\emptyset}(\blacklozenge\varphi)[w]
\end{aligned}$$

□

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 formulas 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. [47, 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.[4]

3 Bisimulation for SML

In this section, we introduce a notion of 'sabotage bisimulation' for SML, a task left open in [34,33]. The invariance results enabled by this new notion of bisimulation (that is, Propositions 2 and 3 below) were independently proven in [9] and [4], albeit in slightly different formulations.

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

3.1 Sabotage bisimulation

Definition 2 (s-bisimulation). *Let* $\mathcal{M}_1 = (W_1, R_1, V_1)$, $\mathcal{M}_2 = (W_2, R_2, V_2)$ *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* $v' \in W_2$ *s.t.* vR_2v' *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* $(\mathcal{M}_1, w)\mathbf{r}(\mathcal{M}_1', w)$*, then there exists* $\mathcal{M}_2'$ *such that* $(\mathcal{M}_2, v)\mathbf{r}(\mathcal{M}_2', v)$ *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* $(\mathcal{M}_2, v)\mathbf{r}(\mathcal{M}_2', v)$*, then there exists* $\mathcal{M}_1'$ *such that* $(\mathcal{M}_1, w)\mathbf{r}(\mathcal{M}_1', w)$ *and* $(\mathcal{M}_1', w)Z(\mathcal{M}_2', v)$*.*

For brevity we write $(\mathcal{M}_1, w) \leftrightarrows_s (\mathcal{M}_2, v)$ *if there exists an s-bisimulation* Z *such that* $(\mathcal{M}_1, w)Z(\mathcal{M}_2, v)$*.*

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.

3.2 Bisimulation and modal equivalence in SML

We first show that s-bisimulation implies SML equivalence.

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

Proof. The proof is by induction on the syntax of φ. Let $(\mathcal{M}_1, w)Z(\mathcal{M}_2, v)$. *Base Case:* The Atom clause of Definition 2 covers the propositional constants. *Induction Step:* The Boolean cases are routine as usual. The Zig$_\Diamond$ and Zag$_\Diamond$ clauses of Definition 2 take care of $\Diamond$-formulas in a standard way familiar from basic modal logic. As for $\blacklozenge$-formulas, assume that $\mathcal{M}_1, w \models \blacklozenge\varphi$. By the semantics of $\blacklozenge$, we have $(\mathcal{M}_1, w)\mathbf{r}(\mathcal{M}_1', w)$ and $\mathcal{M}_1', w \models \varphi$ and, by clause Zig$_\blacklozenge$ of Definition 2, it follows that there exists $\mathcal{M}_2'$ such that $(\mathcal{M}_2, v)\mathbf{r}(\mathcal{M}_2', v)$ 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 in case the models at issue are 'ω-saturated'. To introduce this notion, we need 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 $\mathcal{M}_Y$.[5] In what follows, $\mathbf{x}$ is a finite tuple of variables.

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 $\Gamma(\mathbf{x})$ of $\mathcal{L}^1_Y$-formulas whose finite subsets $\Gamma'(\mathbf{x}) \subseteq \Gamma(\mathbf{x})$ are all realized in $\mathcal{M}_Y$.*

Thus, a model $\mathcal{M}$ is ω-saturated if for any set of formulas $\Gamma(\mathbf{x}, y_1, \ldots, y_n)$ over a finite set of 'running variables' $\mathbf{x}$ and 'parameters' $y_1, ..., y_n$, once some interpretation of the $y_1, \ldots, y_n$ is fixed to, say, $w_1, \ldots, w_n$, and all finite subsets of $\Gamma(\mathbf{x})[w_1, \ldots, w_n]$ are realizable in $\mathcal{M}$, then the whole of $\Gamma(\mathbf{x})[w_1, \ldots, w_n]$ is realizable in $\mathcal{M}$. From a standard 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. However, what we need for SML is slightly stronger, since we need to find pairs of states so as to satisfy a given type involving deletion, as we shall see.

Proposition 3 ($\leftrightsquigarrow_s \subseteq \leftrightarrows_s$). *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 $(\mathcal{M}_1, w_1) \leftrightarrows_s (\mathcal{M}_2, w_2)$.*

Proof. We show that the relation $\leftrightsquigarrow_s$ itself is an s-bisimulation (cf. 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 $\mathrm{Zig}_\Diamond$ and $\mathrm{Zag}_\Diamond$ proceeds as usual for basic modal languages. Next, we prove that the condition $\mathrm{Zig}_{\blacklozenge}$ is satisfied. Assume that $(\mathcal{M}_1, w_1) \leftrightsquigarrow_s (\mathcal{M}_2, w_2)$ and $(\mathcal{M}_1, w_1)\mathbf{r}(\mathcal{M}'_1, w_1)$. We show there must be a model $(\mathcal{M}'_2, w_2)$ such that $(\mathcal{M}_2, w_2)\mathbf{r}(\mathcal{M}'_2, w_2)$ and $(\mathcal{M}'_1, w_1) \leftrightsquigarrow_s (\mathcal{M}'_2, w_2)$. 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:

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

The first equivalence holds by the assumption of sabotage equivalence between $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$. The second follows by Theorem 1 and the third one by Definition 1. From this, by ω-saturation of $\mathcal{M}_2$ with respect to types having pairs of running variables as mentioned above, we can conclude that:

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

[5] For basic facts on ω-saturation we refer the reader to [12, Ch. 2] and [13, Ch. 2]. The first use of the following proof method for modal logic is found in [43].

Using Theorem 1, it follows that there is a pointed model $(\mathcal{M}'_2, w_2)$ with $(\mathcal{M}_2, w_2)\mathbf{r}(\mathcal{M}'_2, w_2)$ and $\mathcal{M}'_2 \models ST_x^{\emptyset}(\mathbb{T}^s(\mathcal{M}'_1, w_1))[w_2]$. Finally, it is immediate that $(\mathcal{M}'_1, w_1) \leftrightsquigarrow_s (\mathcal{M}'_2, w_2)$, which completes the proof of the $\text{Zig}_\blacklozenge$ clause.

In the same way it can be shown that also the condition $\text{Zag}_\blacklozenge$ is satisfied. [6] □

We have thus established a precise match between sabotage modal equivalence and sabotage bisimulation for the special class of ω-saturated models.[7]

4 Characterization of SML by Invariance

We now extend a classical result for basic modal logic. We characterize SML as the one free variable fragment of FOL that 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:

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

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

(a) If $\mathbb{C}^s(\varphi) \models \varphi$, then φ is equivalent to the translation of an $\mathcal{L}^s$-formula.

(b) $\mathbb{C}^s(\varphi) \models \varphi$ – 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 $\models \bigwedge \Gamma \rightarrow \varphi$ for some finite $\Gamma \subset \mathbb{C}^s(\varphi)$. The converse holds by the definition of $\mathbb{C}^s(\varphi)$: $\models \varphi \rightarrow \bigwedge \Gamma$. 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.

[6] One final thing to check is that the new models obtained are still ω-saturated. This requires a small additional argument translating formulas satisfied at points s in the models $\mathcal{M}'$ after the link deletion into formulas satisfied at s in the original $\mathcal{M}$.

[7] Obtaining this result for any two models requires using an infinitary version of SML.

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

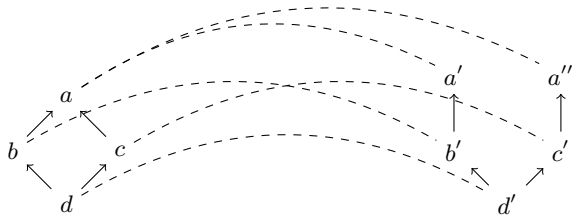

Fig. 2. Two s-bisimilar models (the s-bisimulation runs via the dashed lines). At state d to the left, the property "every two successors share a joint successor" is true. But it fails at point d' in the model to the right.

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 $\Gamma \subseteq ST_x^{\emptyset}(\mathbb{T}^s(\mathcal{M}, w))$. 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 $\Gamma \subseteq ST_x^{\emptyset}(\mathbb{T}^s(\mathcal{M}, w))$.

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 that $(\mathcal{M}, w) \leftrightsquigarrow_s (\mathcal{M}', w')$ as both have the same SML 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, from $\mathcal{M}' \models \varphi[w]$ (by the construction of Σ), we get that $\mathcal{M}'_\omega \models \varphi[w]$. By the assumption that φ is invariant for s-bisimulation and Proposition 3, we get $\mathcal{M}_\omega \models \varphi[w]$ – and then, by elementary extension, $\mathcal{M} \models \varphi[w]$. This completes the proof. □

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.

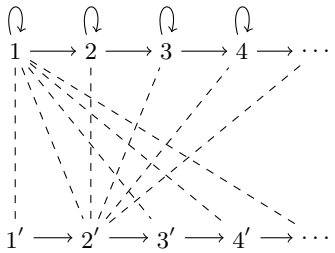

Fig. 3. Two infinite chains $(\mathbb{N}, \geq)$ and $(\mathbb{N}, >)$, with transitive edges omitted in both frames. Only the part of the s-bisimulation relation originating in points 1 and $2'$ is depicted. The remaining edges at other points are positioned similarly.

Caveat. Technically, some care is needed here, because SML works in a universe of changing models. For a sabotage bisimulation, one should draw not only the two base models, but also all sub-models obtainable by one edge deletion, two edge deletions, and so forth. So it is worth spending some time to see why the dashed lines in the relatively simple 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 submodels that can be reached on the left, respectively right, model via exactly i edge deletions.

For the base case, with $(\mathcal{M}_4, s) \in \mathbf{r}^4(\mathcal{M}, d)$ 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: no successors exist, no further deletion can be carried out. For the induction step, first notice that, for any $1 \leq i \leq 4$, for any $(\mathcal{M}_{i-1}, s) \in \mathbf{r}^{i-1}(\mathcal{M}, d)$, there exists $(\mathcal{M}'_{i-1}, s') \in \mathbf{r}^{i-1}(\mathcal{M}', d')$ 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 $(\mathcal{M}_{i-1}, s) \in \mathbf{r}^{i-1}(\mathcal{M}, d)$, there exists $(\mathcal{M}'_{i-1}, s') \in \mathbf{r}^{i-1}(\mathcal{M}', d')$ 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 by $i-1$ deletions on the left, there is a pointed submodel reachable by $i-1$ deletions on the right such that Z is a sabotage bisimulation connecting the two.

A visual inspection of the bisimulation relation depicted in Figure 2 shows that this is indeed the case. Hence confluence is not SML-definable.

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

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 $(\mathcal{M}, 0) \leftrightarrows_s (\mathcal{M}', 0)$. Figure 3 shows this fact by depicting (part of) a relation which is a standard modal bisimulation Z between the two 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, $\mathcal{M} \models xRx[0]$, whereas $\mathcal{M}' \not\models xRx[0]$. The reflexivity property xRx is therefore not definable in SML.

5 Proof Systems 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 dynamic system. Indeed, we have not been able to find a natural Hilbert-style formulation for SML. Therefore, we turn to methods that are easier to formulate staying close to the semantics. This flexibility is provided when we deal with the satisfiability problem by *semantic tableaux* and when we deal with the validity problem by their dual, the *labeled sequent calculi.*

5.1 A Tableau Method for SML

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

Definition 4 (Label, (extended) labeled formula and (relation) term). *Let S be an infinite set whose elements are called* labels. *A* labeled formula *is an expression of the form $x : \varphi$ where x is a label and $\varphi \in \mathcal{L}^s$. A* term *is a pair (x_1, x_2), where $x_1, x_2 \in S$. A* relation term *is an expression of the form $x_1 R x_2$ where $x_1, x_2 \in S$. An* extended label *is an expression of the form x^E where $x \in S$ and E is a finite set of terms. An* extended labeled formula *is an expression of the form $x^E : \varphi$ where x^E is an extended label and $\varphi \in \mathcal{L}^s$.*

[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 $x^{\emptyset} : \varphi$.
2. Repeat the following steps as long as possible:
   (a) *Choose* a branch which is neither closed nor open and choose an extended labeled formula $x^E : \psi$ (or a pair of extended labeled formula $x^E : \psi$ and relation term $x_1 R x_2$) not selected before on this branch.
   (b) *Apply* the appropriate tableau rule of Figure 5 to $x^E : \psi$ (or the pair $x^E : \psi$, $x_1 R x_2$):
      – 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 extended labeled formulas $x^E : p$ and $x^F : \neg p$ (where E and F may possibly be different sets) or two extended labeled formulas $x^E : \varphi$ and $x^E : \neg\varphi$.
   ii. *Label* by $\odot$ (*open*) the (new) branches where there are no more formulas to decompose.

Fig. 4. Construction of a SML tableau.

Labels x represent worlds of modal models, while a relation term $x_1 R x_2$ represents that the pair of worlds represented by (x_1, x_2) belongs to the accessibility relation R. The set of pairs E in x^E represents the pairs of worlds 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 extended labeled formulas 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 formulas 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 x_0, which is accessible from x, may satisfy φ, but also

$$\frac{x^E : \varphi \wedge \psi}{x^E : \varphi \; x^E : \psi}\ \wedge \qquad \frac{x^E : \neg(\varphi \wedge \psi)}{x^E : \neg\varphi \mid x^E : \neg\psi}\ \neg\wedge \qquad \frac{x^E : \neg\neg\varphi}{x^E : \varphi}\ \neg\neg$$

$$\frac{x_1^E : \neg\Diamond\varphi \quad x_1 R x_2}{x_2^E : \neg\varphi}\ \neg\Diamond \qquad \frac{x^E : \neg\blacklozenge\varphi \quad x_1 R x_2}{x^{E\cup\{(x_1,x_2)\}} : \neg\varphi}\ \neg\blacklozenge$$

where $(x_1, x_2) \notin E$ in both rules above.

$$\frac{x^E : \Diamond\varphi}{xRx_1 \; , x_1^E : \varphi \mid \ \ldots \ \mid xRx_n \; , x_n^E : \varphi \mid xRx_0 \; , x_0^E : \varphi}\ \Diamond$$

where $\{x_1, \ldots, x_n\}$ are all the labels occurring in the current branch such that $(x, x_i) \notin E$ for all $i \in \{1, \ldots, n\}$ and x_0 is a 'fresh' label not occurring in the current branch.

$$\frac{x^E : \blacklozenge\varphi}{x_1 R x_1' \; , x^{E\cup\{(x_1,x_1')\}} : \varphi \mid \ \ldots \ \mid x_n R x_n' \; , x^{E\cup\{(x_n,x_n')\}} : \varphi}\ \blacklozenge$$

where $\{(x_1, x_1'), \ldots, (x_n, x_n')\} = (M \times M) \cup \{(x_+, x_{++})\} \setminus E$, with M the set of labels occurring in the current branch to which we add a 'fresh' label x_0, and (x_+, x_{++}) is a pair of 'fresh' and distinct labels.

Fig. 5. Tableau rules for SML.

that one of the worlds/labels x_i of the current model, which have already been introduced, may satisfy φ. In that case we also put an accessibility relation from x to this old world/label x_i. The rule $\blacklozenge$ follows the same kind of resasoning: we may 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.

[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.

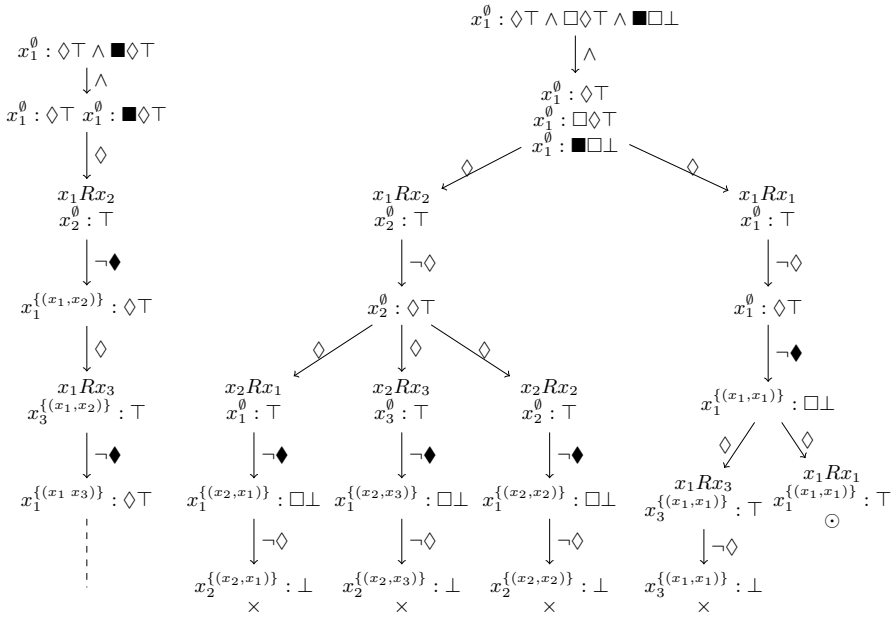

Fig. 6. Two extreme scenarios. An infinite branch arising in the tableau for the formula $\Diamond\top \wedge \blacksquare\Diamond\top$ (*left*). A finite tableau for $\Diamond\top \wedge \Box\Diamond\top \wedge \blacksquare\Box\bot$ (*right*).

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.

Example 6 (Failure of FMP in an extended logic). A somewhat complex failure of FMP is given in [19]. Here we give a more concise formula that uses some extra expressivity resources. Extend the SML language with a converse modality P (for 'past'— we can think of the above $\Diamond$ as 'future') plus a universal modality U over all states. The successive conjuncts of the following formula then say that the accessibility relation is a function, that is one-to-one, but not surjective:

$$U((\Diamond\top \wedge \blacklozenge\Box\bot) \wedge (P\top \rightarrow \blacklozenge\neg P\top) \wedge E\neg P\top). \tag{15}$$

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

5.2 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).*

Soundness Soundness is proved using the notion of *interpretability*. A set of extended labeled formulas L is *interpretable* if there is a Kripke model $\mathcal{M} = (W, R, V)$ and a mapping $f : S \longrightarrow W$ (recall that elements of S are used as labels) such that for all $x^E : \varphi \in L$, we have that $(W, R \setminus f(E), V), f(x) \models \varphi$, where $f(E) = \{(f(x_1), f(x_2)) \mid (x_1, x_2) \in E\}$. 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 extended labeled formulas of some branch is interpretable.*

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

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

Proof (Fact 2). We prove the first fact by induction on the number n of times 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 $\{x : \varphi\}$ and it suffices to define f so that it assigns to x 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. We show that for each rule we can extend the mapping f, which is associated to an interpretable branch, in order to assign world(s) to the new label(s) created by the application of the rule. We show then that 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. In fact, 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 $\Diamond$:* Let the interpretable branch contain an extended labeled formula $x^E : \Diamond\varphi$ (not chosen before when executing 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 $(W, R \setminus f(E), V), f(x) \models \Diamond\varphi$. Hence, there is $w \in W$ such that $f(x)Rw$ and $\mathcal{M}, w \models \varphi$. If w already corresponds to a label x' such that $f(x') = w$, then we are in one of the first n extensions of the interpretable branch. In that case, the map f need not be extended and the label x' 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 x_{n+1} the possible world w: we set $f(x_{n+1}) \triangleq w$.

- *Rule $\blacklozenge$:* Let the interpretable branch contain an extended labeled formula $x^E : \blacklozenge\varphi$ (not chosen before in the execution of the tableau method). Applying Rule $\blacklozenge$ to this interpretable branch, we get n extended branches corresponding to the n elements of $(M \times M) \cup \{(x_+, x_{++})\} \setminus E$. We show that one of them is interpretable. By assumption, we have $(W, R \setminus f(E), V), f(x) \models \blacklozenge\varphi$. Therefore, there is a pair $(w, v) \in R \setminus f(E)$ such that

$(W, R \setminus (f(E) \cup \{(w,v)\}), V), f(x) \models \varphi$. This pair of worlds (w,v) is either of the form $(f(x), f(x'))$, $(f(x), f(x))$, $(f(x), w_0)$, $(w_0, f(x))$, (w_0, w_0) or simply (w_+, w_{++}), for some labels x, x' 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 (x_+, x_{++}) 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 proceed by contraposition. Assume that any extension of the initial interpretable branch closes. Then, any such extension contains an extended labeled formula $x^E : \varphi$ and its negation $x^E : \neg\varphi$ or two extended labeled formulas $x^E : p$ and $x^F : \neg p$. Hence the set of extended labeled formulas of any extended branch is not interpretable. So, by Fact 2, since φ is satisfiable by assumption, this entails that the set of extended labeled formulas of the *initial* branch is also not interpretable, which is impossible by assumption. □

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 extended labeled formulas 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:

- $W = \{x \mid x^E : \varphi \in L \text{ for some } \varphi \in \mathcal{L}^s \text{ and } E \in 2^{S \times S}\}$;
- $R = \{(x_1, x_2) \mid x_1 R x_2 \in T\}$ and
- $V(p) = \{x \in W \mid x^E : p \in L\}$, for all $p \in \mathbf{P}$.

Then, we have the following fact:

Fact 4 *For all extended labeled formulas $x^E : \chi$,*

$$\text{if } x^E : \chi \in L \text{ then } (W, R \setminus E, V), x \models \chi. \tag{16}$$

Proof. We prove Expression (16) 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 $x^E : \varphi \wedge \psi \in L$. Then, by saturation of the tableau rules, we also have that $x^E : \varphi$ and $x^E : \psi$ are in L. Then, by Induction Hypothesis, we must have that $(W, R \setminus E, V), x \models \varphi$ and $(W, R \setminus E, V), x \models \psi$. Hence, we obtain that $(W, R \setminus E, V), x \models \varphi \wedge \psi$.
- $\chi = \Diamond\varphi$: Assume that $x^E : \Diamond\varphi \in L$. Then, by saturation of the tableau rules, there is $xRx' \in T$ such that $(x, x') \notin E$ and $x'^E : \varphi \in L$. Then, by Induction Hypothesis, $(W, R \setminus E, V), x' \models \varphi$ and $(x, x') \in R \setminus E$. Hence, we have that $(W, R \setminus E, V), x \models \Diamond\varphi$.
- $\chi = \blacklozenge\varphi$: Assume that $x^E : \blacklozenge\varphi \in L$. Then, by saturation of the tableau rules, there is $xRx' \in T$ such that $(x, x') \notin E$ and $x^{E \cup \{(x,x')\}} : \varphi \in L$. Then, by Induction Hypothesis, $(W, R \setminus (E \cup \{(x, x')\}), V), x \models \varphi$ and $(x, x') \in R \setminus E$. Hence, we have that $(W, R \setminus E, V), x \models \blacklozenge\varphi$.

- $\chi = \neg p$: Assume that $x^E : \neg p \in L$ and assume towards a contradiction that $(W, R \setminus E, V), x \models p$. Then, by definition of $\mathcal{M}$, there is a set of pairs of labels F such that $x^F : p \in L$. However, if both $x^E : \neg p$ and $x^F : p$ belong to the same branch, the branch cannot be open, which is impossible by assumption.
- $\chi = \neg(\varphi \wedge \psi)$: Assume that $x^E : \neg(\varphi \wedge \psi) \in L$. By saturation of the tableau rules, we have either that $x^E : \neg\varphi \in L$ or $x^E : \neg\psi \in L$. Then, by Induction Hypothesis, we have either that $(W, R \setminus E, V), x \models \neg\varphi$ or $(W, R \setminus E, V), x \models \neg\psi$. In both cases, we obtain that $(W, R \setminus E, V), x \models \neg(\varphi \wedge \psi)$.
- $\chi = \neg\neg\varphi$: Assume that $x^E : \neg\neg\varphi \in L$. Then, by saturation of the tableau rules, we obtain that $x^E : \varphi \in L$. So, by Induction Hypothesis, we have that $(W, R \setminus E, V), x \models \varphi$, and therefore also $(W, R \setminus E, V), x \models \neg\neg\varphi$.
- $\chi = \neg\Diamond\varphi$: Assume that $x^E : \neg\Diamond\varphi \in L$. Then, for all $xRx' \in T$ such that $(x, x') \notin E$, $x'^E : \neg\varphi \in L$ by saturation of the tableau rules. Therefore, by Induction Hypothesis, $(W, R \setminus E, V), x' \models \neg\varphi$, for all $(x, x') \in R \setminus E$. So, $(W, R \setminus E, V), x \models \neg\Diamond\varphi$.
- $\chi = \neg\blacklozenge\varphi$. Assume that $x^E : \neg\blacklozenge\varphi \in L$. Then, by saturation of the rule $\neg\blacklozenge$ of the tableau rules, for all $x_1 R x_2 \in T$ such that $(x_1, x_2) \notin E$, we must have that $x^{E \cup \{(x_1, x_2)\}} : \neg\varphi \in L$. So, by Induction Hypothesis, this entails that $(W, R \setminus (E \cup \{(x_1, x_2)\}), V), x \models \neg\varphi$ for all $(x_1, x_2) \in R \setminus E$. That is, $(W, R \setminus E, V), x \models \neg\blacklozenge\varphi$.

Thus, in particular, since $x^\emptyset : \varphi \in L$ is the root of the tableau for φ, we have from Expression (16) that $\mathcal{M}, x \models \varphi$. Hence, φ is satisfiable.

This proves the completeness of our tableau method.

5.3 A Labeled Sequent Calculus for SML

In classical logic, it is well-known that tableau methods and sequent calculi are interdefinable: proof trees are negated tableau trees turned upside down. Given that our tableau method is based on *labeled* formulas, we provide a *labeled* sequent calculus for SML [20,38]. First, we define the set of *structures* $\mathcal{L}^X$:

$$\mathcal{L}^X \quad : \quad X \quad ::= \quad x : \varphi \quad | \quad xRx \quad | \quad (x, x) \quad | \quad X \,,\, X$$

where $x \in S$ and $\varphi \in \mathcal{L}^s$. We write $x \in X$ when x is a label occurring in X and $X \cup \{x_0\}$ is the set of labels occurring in X to which we add the fresh label x_0. A *sequent* is an expression of the form $X \vdash Y$, $\vdash X$ or $X \vdash$, where $X, Y \in \mathcal{L}^X$.

Theorem 4 (Soundness and completeness). *Let $\varphi \in \mathcal{L}^s$. Then, φ is valid in* SML *if, and only if, $\vdash x : \varphi$ is provable in $\mathsf{L}_{\mathsf{SML}}$ (defined in Figure 7).*

Proof (sketch). Soundness is standard. Completeness follows from Theorem 3 and the following fact: if the tableau for $\neg\varphi$ closes then $\vdash x : \varphi$ is provable in $\mathsf{L}_{\mathsf{SML}}$. Indeed, a closed tableau tree can be transformed into a proof tree in $\mathsf{L}_{\mathsf{SML}}$. Each closed tableau branch is transformed into a branch of the proof tree whose leaf is the axiom $x : p \,, X \vdash x : p$, where X gathers all the terms occuring in the tableau branch. Each tableau rule is transformed into a labeled

$$\frac{}{A\ ,\ X \vdash Y\ ,\ A} \quad A \text{ is either } x:p,\ x_1Rx_2 \text{ or } (x_1,x_2)$$

$$\frac{x:\varphi\ ,\ x:\psi\ ,\ X \vdash Y}{x:\varphi\wedge\psi\ ,\ X \vdash Y}\ \wedge_A \qquad \frac{X \vdash Y\ ,\ x:\varphi \quad X \vdash Y\ ,\ x:\psi}{X \vdash Y\ ,\ x:\varphi\wedge\psi}\ \wedge_K$$

$$\frac{X \vdash Y\ ,\ x:\varphi}{x:\neg\varphi\ ,\ X \vdash Y}\ \neg_A \qquad \frac{x:\varphi\ ,\ X \vdash Y}{X \vdash Y\ ,\ x:\neg\varphi}\ \neg_K$$

$$\frac{x_1Rx_2\ ,\ x_2:\varphi\ ,\ X \vdash Y \quad x_2 \in X \cup \{x_0\}}{x_1:\Diamond\varphi\ ,\ X \vdash Y}\ \Diamond_A \qquad \frac{X \vdash Y\ ,\ x_2:\varphi}{x_1Rx_2\ ,\ X \vdash Y\ ,\ x_1:\Diamond\varphi}\ \Diamond_K$$

$$\frac{x_1Rx_2\ ,\ (x_1,x_2)\ ,\ x:\varphi\ ,\ X \vdash Y \quad (x_1,x_2) \in (X \cup \{x_0\})^2 \cup \{(x_+,x_{++})\}}{x:\blacklozenge\varphi\ ,\ X \vdash Y}\ \blacklozenge_A$$

$$\frac{(x_1,x_2)\ ,\ X \vdash Y\ ,\ x:\varphi}{x_1Rx_2\ ,\ X \vdash Y\ ,\ x:\blacklozenge\varphi}\ \blacklozenge_K \qquad \frac{X\ ,\ X\ ,\ Y \vdash Z}{X\ ,\ Y \vdash Z}\ C_A$$

In all rules, $(x_1,x_2) \notin X$, X, Y can be empty, x_2 is not in the conclusion of $\Diamond_A$.

Fig. 7. Cut-free labeled sequent calculus $\mathsf{L}_{\mathsf{SML}}$ for SML.

sequent calculus rule via the following correspondences: rules $\wedge, \neg\wedge, \neg\neg$ correspond repectively to $\wedge_A, \neg_K \wedge_K \neg_A, \neg_K\neg_A$, and rules $\Diamond, \neg\Diamond, \blacklozenge, \neg\blacklozenge$ correspond respectively to $\Diamond_A, \neg_K\Diamond_K\neg_A, \blacklozenge_A, \neg_K\blacklozenge_A\neg_A$. Moreover, each time the same term is used at several occasion in a tableau branch, we resort in the corresponding proof tree of the sequent calculus $\mathsf{L}_{\mathsf{SML}}$ to the rule of contraction C_A.

Corollary 1. *The cut rule and the weakening rule are admissible in* $\mathsf{L}_{\mathsf{SML}}$.

6 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?

6.1 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 [51]).

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:

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

Behind this result lies a much more general method, first presented in [51]. 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-style logics 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 from 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 $W \times W$.

6.2 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 modalities cannot be pushed through negations the way they are in DEL, since

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

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 [46].

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 property 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 $\langle -p\rangle\Diamond\neg p$ 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[13] —and in that case, stepwiseness would override definability.

6.3 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

[13] Intuitively, deleting points looks like deleting all links arriving or leaving at them.

of links that start at the distinguished point of the current model.[14] Here we provide a few observations to increase familiarity with this system.

Some validities All the principles listed earlier (Formulas (1) to (6)) 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*:*

$$\blacklozenge\Box\bot \to \blacksquare\Box\bot \tag{17}$$

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

$$(\Diamond\Diamond\top \wedge \blacksquare\Box\bot) \to (\Box\varphi \to \varphi) \tag{18}$$

We leave the easy verifications to the reader. In the second formula, the global reading of the deletion modality $\blacksquare$ in the antecedent enforces that the current point is reflexive (cf. formula (14)), 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 $\text{Zig}_\Diamond$ and $\text{Zag}_\Diamond$ clauses of Definition 2 take care of $\Diamond$-formulas in a standard way familiar from basic modal logic. As for $\blacklozenge$-formulas (where $\blacklozenge$ denotes now the *local* variant of sabotage), assume that $\mathcal{M}, w \models \blacklozenge\varphi$. By the semantics of $\blacklozenge$, we have $(\mathcal{M}, w)\mathbf{r}(\mathcal{M}_1, w)$ and $\mathcal{M}_1, w \models \varphi$ where $\mathcal{M}_1$ is obtained via a *local* deletion and, by clause $\text{Zig}_\blacklozenge$ of Definition 2, it follows that $(\mathcal{M}', w')\mathbf{r}(\mathcal{M}'_1, w')$ 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.[15] 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 $\mathcal{M}, w \models \blacklozenge\varphi$ by clause $\text{Zag}_\blacklozenge$ of Definition 2. □

[14] 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.

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

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.

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):[16]

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

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 *and pointed model $(\mathcal{M}, w)$:*

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

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.

6.4 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. [42]).[17] 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.

[16] 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.

[17] In [42] 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 and hybrid logic 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 mechanism for variable binding. In modal terms, we can use devices from hybrid logics [5], such as the universal modality, and especially the 'binder' $\downarrow$, to which we referred also in the previous section. 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. However, 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.[18]

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

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 that we could discuss here: even DEL itself has developed 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. Here, we will just make a few references to further modal and first-order logics of relation change in Section 8, and give a first rough road map.

7 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.

7.1 The sabotage game

The sabotage game was introduced in [44], further studied in [34], and more recently in [24,53], and [47, 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 for more details and motivations, which range from studying the performance of algorithms under adverse circumstances to scenarios in learning theory.

[18] As we observed earlier, 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.

The game is played on a 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 [55]. In the general case, determinacy follows from the Gale-Stewart theorem [23], 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.[19]

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, Demon has a winning strategy. 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 exit 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.[20]

As we will see now, a natural extension of the language of SML can yield the expressivity to characterize winning positions in the sabotage game. It reflects a much more general point about game solutions, developed at length in [47]: game-theoretic equilibria are definable in fixed-point logics for induction and recursion, and often even, in modal fixed-point logics.

7.2 Expressing winning regions in the sabotage game

Our earlier observations pose tests for logics of sabotage games. In particular, Fact 10 suggests that the basic sabotage modal logic SML, possibly with some

[19] Cf. [47] on the background of these results in current logics of games.

[20] 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 multi-relational frames, as in [34].

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:

$$U((\neg \mathtt{goal} \wedge \Diamond\top) \rightarrow \blacklozenge\Box\neg \mathtt{goal}) \tag{19}$$

with U the universal modality and $\mathtt{goal}$ a nominal for the goal point.

Next, for arbitrary goal regions (singletons or not) defined by the formula $\mathtt{goal}$, the original paper [44] observes how the existence of winning strategies for Traveler can be expressed in a PDL-format over SML. At least on finite graphs, the winning positions for Traveler can be defined using an operator for finite iteration, by the formula:

$$(\Diamond\top \wedge \blacksquare\Diamond)^{*}\mathtt{goal} \tag{20}$$

capturing the infinitary disjunction $\mathtt{goal} \vee (\Diamond\top \wedge \blacksquare\Diamond\mathtt{goal}) \vee (\Diamond\top \wedge \blacksquare\Diamond(\Diamond\top \wedge \blacksquare\Diamond\mathtt{goal}))\ldots$ and so forth. Notice that the conjunct $\Diamond\top$ is required to rule out as a winning position the situation in which there are no edges to be deleted in the model but Traveler is not in a goal state. This characterization of Traveler's winning positions, however, does not work on infinite graphs, and in general, a more expressive language is needed.

Let us consider a μ-calculus language μSML in order to define winning positions for Traveler (and Demon) with formulas whose syntax matches the chosen definition of winning positions. Let $\mathbf{P}$ be a set of propositional atoms. Formulas of the sabotage modal language $\mathcal{L}^{\mu}$ have 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. Here, just as ordinary modalities, sabotage modalities do not affect positive or negative occurrence.

Fixpoint operators allow us to generalize Formula (20) and characterize the set of Traveler's winning positions in its full generality, via the formula:

$$\mu p.(\mathtt{goal} \vee (\Diamond\top \wedge \blacksquare\Diamond p)) \tag{21}$$

We have not given a precise semantics for μSML yet, but intuitively the formula makes sense for our game. Traveler is in a region that includes the goal points or to which she can return no matter what Demon deletes.

By the determinacy of the sabotage game, the dual formula of (21) should then define the winning region for Demon, which is therefore the largest fixpoint of the equation $p \leftrightarrow \neg\mathtt{goal} \wedge (\Box\bot \vee \blacklozenge\Box p)$. Viewed in this light, the earlier Formula (19) for Demon stated that $\neg\mathtt{goal}$ is such greatest fixpoint for singleton goals.

Notice how the very shape of the above fixed-point formula mirrors the precise winning conventions of the game for the players. Clearly, this is a general and flexible feature of logical syntax, to which we will return later on.

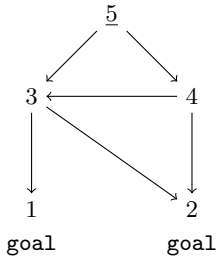

Fig. 8. Graph on which the sabotage game is played in Example 7.

7.3 The sabotage μ-calculus: μSML

To make the preceding precise, we need to specify the semantics of μSML. This may seem obvious: one copies the standard definitions for the μ-calculus[21] and adds a clause for $\blacksquare$, noting that this new operator does not affect positive occurrence. The logic resulting from this obvious extension of the μ-calculus semantics has indeed been studied in [40]. However, this does not yield the intended semantics for analyzing our games, as shown by the following example.

Example 7. Consider a graph as depicted concretely in Figure 8 with vertices $W = \{1, 2, 3, 4, 5\}$, and directed edges $R = \{(3, 1), (3, 2), (4, 3), (4, 2), (5, 4), (5, 3)\}$, whose goal region `goal` is $\{1, 2\}$. Computing, bottom up, the successive approximation stages for the function denoted by $\texttt{goal} \vee (\Diamond\top \wedge \blacksquare\Diamond p)$ yields this sequence of subsets of the graph:

$$\emptyset, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}$$

which ends in the fixpoint $\{1, 2, 3, 4, 5\}$. The reason is that, as the set grows, further points get at least two successors inside it, so one link deletion can never prevent them from getting there. However, this outcome is wrong, since point 5 is not a winning position for Traveler in the graph game as we have defined it. Demon can win by starting with cutting the link $(3, 1)$ and then continuing to cut appropriately as Traveler tries to move to a goal point.

What has gone wrong here? The problem is the model-changing character of the modality $\blacksquare$: the same formula may get different truth values at the same points in the two models before and after a link deletion. Using a mere set of points as the current approximation does not keep track of this in the right way. An improved semantics that does work will only be sketched here.

One direct approach defines *syntactic* approximation formulas φ^α for each ordinal α, in an infinitary version of sabotage modal logic SML. In any given

[21] The standard clause is $\mathcal{M}, w \models \mu p.\varphi(p)$ iff $w \in \bigcap \left\{ X \in 2^W \mid ||\varphi||_{\mathcal{M}[p:=X]} \subseteq X \right\}$, where $||\varphi||_{\mathcal{M}[p:=X]}$ is the truth-set of φ with $V(p)$ set to be X.

model, this sequence will stabilize at some stage after which all later formulas in the sequence have the same truth-set, thereby computing the smallest fixpoint. Of course, the formula at that stage may depend on the size of the model.[22]

But our preferred approach is *semantic*, changing the relevant universe of evaluation to better fit our intuitions, and in line with the notion of sabotage bisimulation (Definition 2). Instead of working on just the original graph, we consider a dynamic model $\mathbf{M}$ consisting of this graph and all those obtainable from it by edge deletions. We take the points of this new model to be pointed models $(\mathcal{M}, s)$, and let the modality $\Diamond$ access the relation between $(\mathcal{M}, s)$, $(\mathcal{M}, t)$ with $(s, t) \in R^{\mathcal{M}}$, and $\blacklozenge$ access the sabotage relation $\mathbf{r}$ between $(\mathcal{M}, s)$ and $(\mathcal{M}^-, s)$ where $\mathcal{M}^-$ arises from $\mathcal{M}$ by deleting one directed edge.[23] Now we can just compute on this model with the original formula, and get an increasing family of sets that yields the same results as the syntactic procedure.

We leave it to the reader to recompute the earlier example in terms of pointed models, and check that we now get the intended interpretation.

The system obtained in this way is not a simple extension of the μ-calculus. It is still easy to show that $\mu\mathsf{SML}$ formulas are invariant under sabotage bisimulation. It is less clear, however, how to extend the earlier standard translation ST. It might seem to extend immediately to a translation into $\mathsf{FO}(\mathsf{LFP})$, the extension of first-order logic with fixed-point operations, by taking modal μp to first-order $\mu P, x$. But this fails to capture evaluation with different truth values across models. Finally, there is a major issue of complexity. This can be illustrated with the logic PAL^* of public announcement logic with finitely iterated announcements. Its formulas $\langle(!\varphi)^*\rangle\psi$ may be seen as solutions for a fixpoint formula $\mu p.\psi \vee \langle!\varphi\rangle p$ computed in our sense across different models. But [36] shows that, unlike the μ-calculus, this system is undecidable, in fact its satisfiability problem is Π^1_1-complete. We leave these matters for future investigation.

Coda Our semantic view with pointed models as the new indices of evaluation also has some independent interest, in that it suggests a natural generalization. A universe need not be a full 'standard model' with all possible variants, say some initial model plus the family of all its sub-models. We could also allow only 'admissible models' in the universe according to some criterion, giving us much more variety in patterns for the model shifts.[24] On such more general models, the complexity of our logical systems might well decrease.

---

[22] One might think that this is the same as working with just sets, since in the standard μ-calculus, 'call by name' and 'call by value' are equivalent in the approximation procedure. But this only holds in one fixed model, not across models where, as we saw in the above example, truth values can differ.

[23] Essentially, we are now in a bimodal model of a special sort, where the atomic `goal` formula holds for the same points across different sub-graphs.

[24] Such universes are studied in dynamic-epistemic logic as 'protocol models', cf. [49].

7.4 General graph games

The general background for the topic of this section is that of graph games. To put the sabotage game in perspective, we briefly consider two more such games.

The travel game For a start, here is a simple game that leaves a given graph unchanged. In the *Travel Game*, Demon picks an outgoing edge of the current point, Traveler has to pick an edge to move from there, and so forth. This time, Traveler wins if Demon cannot issue a challenge (*i.e.*, we are at an endpoint that is Demon's turn), or if Traveler can keep moving forever. This game has been applied widely to model-checking or other logical tasks (cf. [45]), given the great freedom in what we can define to be a graph for the task at hand. The game is determined, and winning positions can be described simply in the modal μ-calculus. In particular, those for Traveler are given by the greatest fixed-point formula

$$\nu p.\Box\Diamond p \tag{22}$$

The modal μ-calculus is a high-level theory of this, and similar graph games.

In contrast with the above, the *Sabotage Game* changes the underlying graph drastically.[25] Accordingly, winning conditions required using a μ-calculus extended with sabotage modalities (cf. Section 7.3).

The poison game Finally, consider a game that does not change graph links, but that changes the 'annotation' of a graph: *i.e.*, unary properties of its nodes. The *Poison Game* of [17] has Demon issuing a challenge as in the Travel Game, while Traveler has to respond. However, now Demon 'poisons' each node that he visits, making it inaccessible to Traveler: she loses if she visits a poisoned node. Traveler wins this game if she can keep going. In the game of the cited paper, the first move is for Traveler, who picks any node as a starting point.

This game captures an important graph-theoretic notion. A 'local kernel' is a non-empty set of nodes that has no internal edges (a so-called independent set), while for every edge leading outside the set, there is a follow-up edge returning into the set.[26] It can be proved relatively easily that, on finite branching and upward well founded graphs, there is a local kernel if and only if Traveler has a winning strategy.

But the Poison Game also has independent interest, and describing winning positions for Traveler suggests the following definition for the winning positions of the Traveler. Consider a μ-calculus with a special predicate P for 'poisoned',

[25] One might think this reduces to a Travel Game over a 'supergraph' of ordinary graphs related by sabotage steps. But, given the complexity results we cited for sabotage modal logic, this translation necessarily blows up computational complexity.

[26] In modal terms, p defines a semi-kernel of the underlying frame, if and only if the formula $Ep \wedge U(p \rightarrow (\Box(\neg p \wedge \Box p))$ holds in the model, with E the global existential modality, and U the universal modality.

and an added modality $[+P]\varphi$ which says that φ holds at the current point s of the model after this has been poisoned. That is, the operation $+P$ changes the current interpretation for P to add the singleton s.

$$\nu q.\neg P \wedge \Box[+P]\Diamond q \tag{23}$$

Intuitively, this language with a dynamic poison modality seems intermediate between the pure μ-calculus and the sabotage μ-calculus.[27] The poison modality involves only what is called 'factual change' of propositional valuations in dynamic-epistemic logic [52]. There are recursion axioms for this modality pushing it through all other operators, turning formulas into equivalent static ones. Still, even though the modal μ-calculus is closed under DEL product update [50], we believe that our caveat for μSML applies also to the a μ-calculus extended with the poison modality, and a new semantics is needed.

8 Related Work

8.1 Other recent studies of sabotage logic

As we have noted right at the start, the complexity and model-theory of SML has been studied extensively in a recent series of works [1,3,2,4] and [19]. In particular [19,4] 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 [34,33]. Our paper contributes further results—specifically, a charaterization theorem—to this specific line of research on the model-theory of SML.

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 $\Diamond$, 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.

8.2 Argumentation games

Abstract argumentation theory [18] studies criteria for tenability of positions in argumentation. Its key structures are directed 'attack graphs' (frames) $G = (W, R)$ where W is a non-empty set of arguments and R is a binary relation aRb interpreted intuitively as saying that argument b attacks (*is stronger than*)

[27] For some results on basic modal logics of poisoning, see also [35].

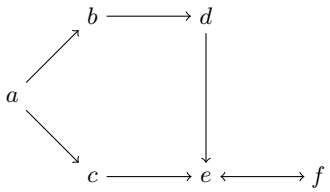

Fig. 9. An attack graph.

X is conflict-free	iff $\nexists a, b \in X$ s.t. aRb
X is self-defended	iff $\forall a, b$ if $a \in X$ and aRb then $\exists c \in X$ s.t. bRc
X is a complete set	iff X is conflict-free and $X = \{a \mid \forall b, \text{if } aRb \text{ then } \exists c \in X \text{ and } bRc\}$
X is the grounded set	iff X is the smallest complete set

Table 1. Some key notions of abstract argumentation theory from [18].

argument a.[28] All graphs illustrated in this paper can therefore be interpreted as attack graphs, but we provide a further example in Figure 9.

A basic question is: given an attack graph, which set of arguments can be rationally upheld? Answers give rise to different criteria for the 'quality' of sets of arguments. We list some basic ones in Table 1. As shown in a series of works [26,27,28,22], all these notions have a modal character, making modal (fixpoint) logics a natural tool for argumentation theory. In particular,

$$\mu p.\Box\Diamond p \tag{24}$$

defines the notion of grounded set [26,27,22], that is, the set of arguments that can always be ultimately defended by a set of unattacked arguments. This intuition is naturally illustrated by the stages of approximation of Formula (24): the empty set; the set of arguments whose attackers are in turn attacked by arguments in the empty set, that is, the unattacked arguments; and so forth. Formula (24) defines the winning positions in a game for Traveler (the 'Proponent' in argumentation theory) who wins if, and only if, she reaches an argument that is unattacked, while Demon (the 'Opponent') is out of counter-arguments.[29]

Games played on attack graphs occur in the argumentation literature as 'dialectical' procedures for 'static' structural criteria of acceptability of arguments

[28] In the literature on abstract argumentation edges are interpreted in the opposite direction: aRb represents that argument a *attacks* argument b.

[29] It may be instructive to compare Formulas (22) and (24). From an argumentation-theoretic point of view, the second imposes a more stringent winning requirement on Traveler, forcing the Demon to get stuck, while the first allows Traveler to simply 'keep going'.

(see [37] for an overview). For instance, a close relative of our earlier poison game is the game for 'credulous admissibility' in [54]. Traveler selects attacks to the last argument selected by Demon and, as in the poison game, cannot select any argument previously selected by Demon. Demon plays by selecting attacks to *any* argument previously selected by Traveler, being allowed to backtrack. Demon starts. Traveler wins if she does not run out of counter-arguments. It can be shown [54] that Traveler has a winning strategy if, and only if, the initial argument is included in a conflict-free and self-defended set (that is, a semi-kernel).

Again, suitably enriched fixpoint logics can capture such games. For instance, using the notation of Section 7.4, Traveler's winning positions are defined by

$$\nu q.((V \wedge \neg P) \wedge U(V \rightarrow \Box[+P]\Diamond[+V]q)) \tag{25}$$

with U the universal modality, P the 'poison' predicate, while V keeps track of the points visited by Traveler. A modal logic of graph games as envisioned in our paper may well have significant applications to abstract argumentation theory.

8.3 Landscape of logics for model change

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 than what we have covered here. Besides the above mentioned series of work [1,3,2,4], relevant systems include local and global graph modifiers [8], dynamic epistemic modifiers [11,10], dynamic modal logic DML [15], arrow logic [31], logics of copy and remove [6], the logic of preference upgrade [51], and general dynamic dynamic logic [25].[30]

Instead of discussing the above terrain in detail, we give a small 'expressiveness map' based on what can readily be observed, and on what authors have claimed.[31] In Figure 10, an arrow $L_1 \rightarrow L_2$ 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. [31]) and preference upgrade logic (cf. [51]). 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].[32] From the cited publications,

[30] Also relevant here is [21] on 'reactive logics', where the accessibility relation of modal models is changed during the interpretation process of formulas.

[31] 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, cf. [25], which is equivalent to PDL. Also, we were unable to fit in the logics of copy and remove.

[32] As observed earlier (footnote 14) the system of 'local' sabotage from [1,3] is subtly different from the one we considered in this paper.

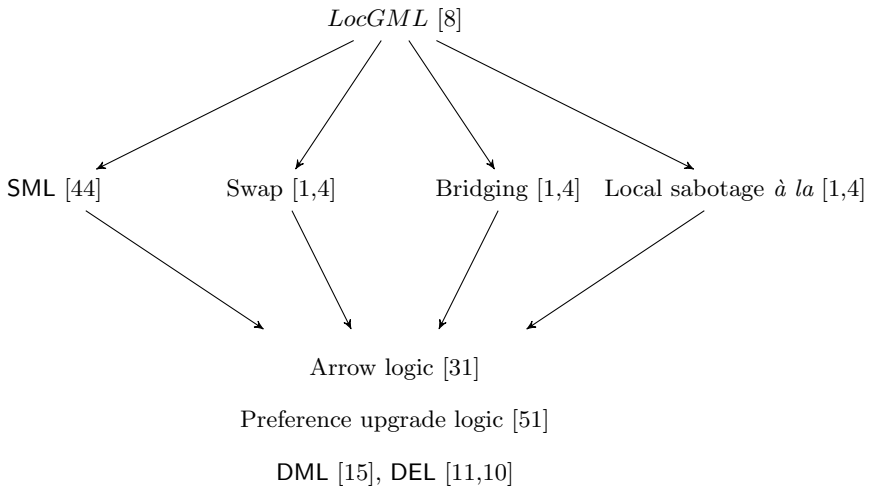

Fig. 10. Expressiveness map of some dynamic modal logics for graph modifiers.

we know that these mid-level systems are incomparable in terms of expressivity, that they are also strictly more expressive than the basic modal logic (cf. [4]) 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 [8] 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.

A similar map would make sense for fixed-point extensions of the logics in this paper. In Section 7, we used the modal μ-calculus, and extensions with DEL-style dynamic modalities. Another natural extension given our comparison of sabotage logics and dynamic-epistemic logics is a fixed-point logic $\mu(\downarrow, U)$ containing the base modalities plus the universal modality and the hybrid binders. This provides the expressive power needed to deal with local sabotage and other dynamic operations that have embedded references to earlier points of evaluation. The literature on graph change also contains other such systems,[33] many of them in between the modal μ-calculus and the logic LFP(FO), first-order logic with fixed-point operators. The latter system is Π^1_1-complete, but it forms a natural limit for many fixpoint logics. However, we have also introduced a sabotage μ-calculus that seemed a much more radical expressive extension. Even lacking a road map like our earlier one, we believe that just an awareness of this realm

[33] For other interesting extended μ-calculi, see, for instance, [40] and [41].

beyond first-order logic poses interesting challenges, such as exploring the reach of current automata-theoretic methods for fixpoint logics.

Of course, comparing logics by mere expressiveness 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 equivalences are all up for further investigation 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.[34]

9 Conclusions and Future Work

In this paper, we have dusted off sabotage modal logic, and looked at its broader current relevance. We gave a first-order translation for this logic, proved a novel characterization theorem in terms of a new notion of sabotage bisimulation, and introduced a sound and complete tableau system and sequent calculus.

Still, SML remains an under-investigated system. 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 is modal correspondence theory, whose techniques extend into higher-order logic [43]. In particular, can the Sahlqvist theorem be generalized to SML? Also, in this paper, we arrived at natural variations of SML, such as its version with only local deletions, whose decidability is open.[35] Next, given the apparent complexity of the system, it makes sense to look at natural fragments, such as the closed formulas that formed many of our examples, or the formulas of sabotage depth 1. Alternatively, in line with a suggestion in Section 8.3, we could look at 'protocol models' [49], that restrict the sequences of available link deletions between graphs, giving rise to bimodal models with an 'internal' accessibility relation R for $\Diamond$ and an 'external' relation $\mathbf{r}$ for $\blacklozenge$. Finally, we also found SML-inspired new versions of dynamic-epistemic logics, such as stepwise PAL, whose properties seem unknown.

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 DEL world. We see our way of looking as the start of something more ambitious, finding a better map of the

[34] 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' [7], based on display-style substructural proof theory.

[35] Other natural variations include multi-modal SML with indices for the relations to be cut, [34,33], or logics of adding links, like the aforementioned modal logics of bridging [4]. In this connection, note that adding links to R is the same as deleting links from the complement $-R$, so there are connection laws to be had.

whole field of logics for model change, and the major parameters that determine their complexity. 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.

Next, returning to the games that motivated SML in the first place, we discussed sabotage μ-calculus as a system for defining strategic powers of players, and as a testing ground for introducing model-changing modalities that can affect the process of approximation leading to the usual smallest or greatest fixed-points. The system we define seems a very natural object of study in its own right. And as we have seen in Section 8, it suggests broader issues about the landscape of further fixed-point logics for model change.

But more importantly than just studying logics, the sabotage game reminds us of a general issue of *gamification*: the use of newly designed games to model significant computational or social scenarios, as argued for in [47].

Sabotage games model real dynamic scenarios unfolding over time, beyond one-step model changes. There are many natural variations on the original scenario in [44] —and these may serve as pilots for more general games played over networks, such as the argumentation games of [26,22] discussed earlier, or the social network games of [48] and [14].[36] Such network games need not be about 'sabotage' at all, and the general area becomes that of games that may change their own playground, for which this paper has provided some logical tools.

Acknowledgments 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 wish to thank Valentin Goranko for insightful discussions on the topic of this paper, Sebastian Enqvist for improvements to one of our proofs, and the reviewers of this special issue for their various helpful comments.

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.

[36] Work in these areas also shows that gamification is not unique. For instance, the μ-calculus formulas that we have used to describe games themselves have standard evaluation games, and the connection between games, describing formulas, and evaluation games for the latter is often an intriguing one.

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 Ulrich Kohlenbach, Pablo Barceló, and Ruy de Queiroz, editors, *Logic, Language, Information, and Computation: 21st International Workshop, WoLLIC 2014, Valparaíso, Chile, September 1-4, 2014. Proceedings*, pages 51–65, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.
7. G. Aucher. Displaying updates in logic. *Journal of Logic and Computation*, 26(6):1865–1912, 2016.
8. 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.
9. G. Aucher, J. van Benthem, and D. Grossi. Sabotage modal logic: Some model and proof theoretic aspects. In *Proceedings of LORI'15*, 2015.
10. A. Baltag and L. Moss. Logics for epistemic programs. *Synthese*, 139(2):165–224, 2004.
11. A. Baltag, L. Moss, and L. 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. G. 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. P. Duchet and H. Meyniel. Kernels in directed graphs: A poison game. *Discrete Mathematics*, 115:273–276, 1993.
18. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.
19. R. Fervari. *Relation-Changing Modal Operators*. PhD thesis, Universidad Nacional de Córdoba, Facultad de Matemática, Astronomía y Física, 2013.
20. D. Gabbay. *Labelled Deductive Systems*. Oxford University Press, 1996.
21. D. Gabbay. *Reactive Kripke Semantics*. Springer, 2013.
22. 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.
23. 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.
24. 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'09*, number 5834 in LNAI, 2009.

25. P. Girard, J. Seligman, and F. Liu. General dynamic dynamic logic. In T. Bolander, T. Braüner, S. Ghilardi, and L. 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.
26. 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.
27. 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.
28. D. Grossi. Fixpoints and iterated updates in abstract argumentation. In *Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR 2012)*, 2012.
29. S. Gruener, F. Radmacher, and W. Thomas. Connectivity games over dynamic networks. *Theoretical Computer Science*, 498:46–65, 2013.
30. 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.
31. B. Kooi and B. Renne. Arrow update logic. *Review of Symbolic Logic*, 4(4), 2011.
32. F. Liu, J. Seligman, and P. Girard. Logical dynamics of belief change in the community. *Synthese*, 191(11):2403–2431, 2014.
33. 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.
34. C. Löding and P. Rohde. Solving the sabotage game is PSPACE-hard. Technical report, Department of Computer Science RWTH Aachen, 2003.
35. C. Mierzewski and F. Zaffora Blando. The modal logic(s) of poison games. Department of Philosophy, University of Stanford.
36. J. Miller and L. Moss. The undecidability of iterated modal relativization. *Studia Logica*, 79(3):373–407, 2005.
37. S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In I. Rahwan and G. Simari, editors, *Argumentation in Artificial Intelligence*, pages 105–132. Springer, 2009.
38. Sara Negri. Proof analysis in modal logic. *J. Philosophical Logic*, 34(5-6):507–544, 2005.
39. 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.
40. P. Rohde. On the mu-calculus augmented with sabotage. In *Foundations of Software Science and Computation Structures, 7th International Conference, FOSSACS 2004, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2004, Barcelona, Spain, March 29 - April 2, 2004, Proceedings*, volume 3921 of *LNCS*, pages 142–156, 2006.
41. U. Sattler and M. Vardi. The hybrid mu-calculus. In R. Goré, A. Leitsch, and T. Nipkow, editors, *Proceedings of IJCAR 2001*, volume 2083 of *LNAI*, pages 76–91. Springer, 2001.

42. 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.
43. J. van Benthem. *Modal Logic and Classical Logic*. Monographs in Philosophical Logic and Formal Linguistics. Bibliopolis, 1983.
44. 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.
45. J. van Benthem. *Modal Logic for Open Minds*. CSLI Publications, 2010.
46. J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2011.
47. J. van Benthem. *Logic in Games*. MIT Press, 2013.
48. 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.
49. J. van Benthem, J. Gerbrandy, H. Tomohiro, and E. Pacuit. Merging frameworks for interaction. *Journal of Philosophical Logic*, 38(5):491–526, 2009.
50. J. van Benthem and D. Ikegami. Modal fixed-point logic and changing models. In *Pillars of Computer Science, Essays Dedicated to Boris (Boaz) Trakhtenbrot on the Occasion of His 85th Birthday*, volume 4800 of *LNCS*, pages 146–165. Springer, 2008.
51. J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, 17(2), 2007.
52. J. van Benthem, J. van Eijck, and B. Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, 2006.
53. F. Velázquez-Quesada. *Small Steps in the Dynamics of Information*. PhD thesis, ILLC, University of Amsterdam, 2011.
54. G. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA'00)*, LNAI, pages 239–253. Springer, 2000.
55. 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.