

# BISIMULATIONS FOR KNOWING HOW LOGICS

RAUL FERVARI

Universidad Nacional de Córdoba and CONICET

FERNANDO R. VELÁZQUEZ-QUESADA

Institute for Logic, Language and Computation, Universiteit van Amsterdam  
and

YANJING WANG

Institute for Foreign Philosophy and Department of Philosophy, Peking University

**Abstract.** As a new type of epistemic logics, the *logics of knowing how* capture the high-level epistemic reasoning about the knowledge of various plans to achieve certain goals. Existing work on these logics focuses on axiomatizations; this paper makes the first study of their model theoretical properties. It does so by introducing suitable notions of bisimulation for a family of five *knowing how* logics based on different notions of plans. As an application, we study and compare the expressive power of these logics.

**§1. Introduction.** Standard epistemic logic [18, 33] studies the reasoning patterns of *knowing that*: the knowledge an agent might have about the truth-value of certain propositions. It has been successfully applied to various fields besides philosophy, such as theoretical computer science, game theory, and AI (see, e.g., [18, 41, 56]). However, *knowing that* is not the only interesting expression of knowledge an agent might use. One can look, for example, at the closely-related *knowing whether*, which indicates that the agent knows the truth-value of a proposition without indicating if she knows that the proposition is true or if she knows that the proposition is false [19, 30]. There is also the drastically different *knowing why*, concerning the reasons/arguments/evidence/justifications the agent has for claiming that a certain proposition is the case [6, 64]. There are also several forms of *knowing what*, discussing variables, their possible values, their possible dependencies, and knowledge about the latter two (e.g., [8, 26, 57]). In [61] Wang presents a survey of logical approaches to these forms of knowledge.

Another important kind of knowledge is expressed by *knowing how*. This form of knowledge is particularly important not only in philosophy [20], but also in applied sciences (see the surveys [3, 23]). In particular, it is important in automated planning and strategic reasoning within AI. For the first, deciding whether an agent knows how to achieve a certain goal is essentially asking whether she has at her disposal a sequence of actions (that is, a strategy or a plan) that will always reach the intended target. For

---

Received: August 23, 2019.

2020 *Mathematics Subject Classification*: 03B42, 03B45, 03B70, 68T27, 68Q85.

*Key words and phrases*: epistemic logic, knowing how, bisimulation.

the second, in multi-agent interaction, (knowing) whether another agent has the ability to guarantee a certain outcome is crucial when deciding our own course of action.

Interestingly, *knowing how* cannot be captured by a simple combination of *knowing that* and ability, as shown in [31, 36]. This is because a philosophically grounded formulation of *knowing how to achieve*  $\varphi$  requires a *de re* reading: there exists a plan  $\sigma$  such that the agent knows that executing  $\sigma$  guarantees  $\varphi$  [53]. This differs from the *de dicto* reading which, combining *knowing that* and ability, would yield ‘*the agent knows that there is a method whose execution guarantees*  $\varphi$ ’. This reading requires for the agent to have a  $\varphi$ -guaranteeing method, but it does not force her to know *which* the method is. Equally important is the fact that a proper notion of *knowing how* may take a *global* perspective: the given strategy should allow the agent to achieve the goal *in every scenario satisfying a given initial condition*, in order to assure that the success is truly a matter of ability, and not a matter of luck.

Based on this reading, Wang proposed and studied a modal logic featuring a simple goal-directed *knowing how* operator Kh [59, 62], of which variants have been studied in [37, 38, 39]. At the semantic level, these logics use relational models, which in this context can be viewed as a description of the abilities of the agent (hence their common alternative name of *ability maps*). From these structures, one can extract the agent’s knowledge-how.

Despite using standard modal logic tools [12], the proposed logics are not normal, and actually they should not be: intuitively, knowing how to get drunk and knowing how to prove a theorem do not imply knowing how to prove the theorem while drunk, in contrast with the valid normal modal logic formula  $(\Box\varphi \wedge \Box\psi) \rightarrow \Box(\varphi \wedge \psi)$ . Moreover, fulfilling the aforementioned requirement, the provided *knowing how* operators are global, that is, the truth value of a *knowing how* formula *does not* depend on the point of evaluation (in fact, they can define the global universal modality discussed in [25]). Finally, these logics have characteristic axioms capturing the composition of plans/strategies.

A potential drawback of the mentioned proposals is that, while the semantics of Kh is based on linear plans, the languages do not have a traditional *knowing that* modality K; thus, the logics do not capture notions as knowledge-based plans or conditional plans. These notions are important in certain scenarios, such as planning problems in AI, where initial uncertainty and nondeterministic actions are present. As a consequence, [21] introduced a logic for *strategically knowing how*, which combines the discussed form of *knowing how* with a standard *knowing that*. In this approach, the ‘global perspective’ is left out, and both modalities are interpreted locally. This is required in order to properly capture the interaction between *knowing how* and the typically local *knowing that* operator. Even more, instead of linear plans, the *knowing how* modality operates now over branching plans, so the agent can make use of the knowledge acquired during the plan’s execution. Another approach to strategy-based knowing how came from the coalition logic tradition, within a multi-agent setting, but with single-step strategies [43, 44, 46, 47] inspired by (earlier versions of) [1]. All these aspects, together with the diversity of proposals, are sufficient to make the family of *knowing how* logics interesting not only for logicians but also for philosophers and computer scientists.

The work of the mentioned proposals focuses mostly on axiomatizations. This paper makes the first attempt to study the model theoretical aspects of these know-how logics, focusing in particular in the study of appropriate notions of

*bisimulation*.<sup>1</sup> A bisimulation is a relation between the states of two models, providing a semantic characterization of the distinguishing power of the given language while also formalizing the idea of behavioral equivalence. More precisely, a bisimulation relates states from two models when they contain the same propositional and relational information, in the respective sense of having both the same labelling as well as ‘matching’ successors. (For a more detailed technical account of bisimulation in modal logic and, in general, notions of equivalence between structures, see e.g., [24, 34].)

Bisimulations are useful for theoretical results, as helping in characterizing the expressive power of logical languages (e.g., the van Benthem characterization theorem for modal logic; [12]). But they also have more practical applications [52]. For example, it is well-known that the expressivity and the complexity of a language are somehow related. In many cases, the high expressivity of the language gives us a hint of the high computational complexity of its inference tasks. Thus, a language’s expressivity provides a guideline on its complexity profile, which can be used to decide which language is adequate for the problem at hand. Bisimulations also have applications in AI planning, where the size of the domain (or, from a logical perspective, the size of the model) is crucial. Here, bisimulations can be used to obtain smaller models by means of a *bisimulation contraction*, which produces a minimal model that inherits all the observable logical properties from the original one. In fact, in epistemic planning, there are already proposals that rely on notions of bisimulation in order to reduce a model or to show some technical result, with [14] being an example. Bisimulations can also be used to *add* states efficiently, as done in [48] when computing the *Merge and Shrink abstraction*, a method for heuristic design in AI planning. These examples show how notions of bisimulations for *knowing how* frameworks can be useful in planning scenarios.

In all the notions of bisimulation that this paper discusses, we need to overcome one technical difficulty: the imbalance between weak languages (a single *knowing how* operator, indicating only whether the agent can guarantee the satisfaction of  $\varphi$  when starting in situations satisfying  $\psi$ ) and rich models (the relational models, indicating not only which actions are available at each state of the system, but also which states are reached after each action takes place). To this end, the notions of bisimulation introduced in this proposal work by abstracting from the details provided by the model. More precisely, they are essentially based on a ‘forcing relation’ that takes the agent from a given set of states  $U$  to those states she can reach (from  $U$ ) by the execution of some plan/strategy.

This paper presents bisimulations for several *knowing how* logics. The first one, discussed in detail, is a bisimulation for the basic *knowing how* logic of [62]. The paper also provides bisimulations for two of its variations, namely the *knowing how* with intermediate constraints of [38] (which additionally asks for the chosen strategy to go only through ‘appropriate’ states), and the weakly *knowing how* of [37] (which weakens the requirement on the required strategy by allowing ‘incomplete’ runs). Bisimulations for logics dealing with knowledge-based plans are also reviewed. We discuss in detail the case of the *strategically knowing how* logic of [21], and we also establish connections with a single-agent variant of the logic introduced in [46]. Other proposals dealing with an agent’s abilities using different kinds of structures have been explored in the

---

<sup>1</sup> A preliminary version of this paper was first presented at SR2017 [22].

Table 1. *A summary of the five know-how logics discussed in this manuscript.*

	Plan	Global/local	Know-that	Executability	Int-cons.	Reference
$\mathcal{L}_{Kh}$	Linear	Global	No	Strong	No	[59]
$\mathcal{L}_{Khm}$	Linear	Global	No	Strong	Yes	[38]
$\mathcal{L}_{Khw}$	Linear	Global	No	Weak	No	[37]
$\mathcal{L}_{Khs}$	Branching	Local	Yes	Strong	No	[21]
$\mathcal{L}_H$	One-step	Local	Yes	Strong	No	[46]

The classification is based on features of the semantics of the know-how operator, to be made clear in the later sections: whether it is based on linear or branching plan, whether the modality is global or local, whether the language also includes know-that operator, whether it requires strong executability, and whether it involves intermediate constraints (int-cons. column).

literature. Examples include frameworks in the STIT tradition (e.g., [17, 32, 35]) and approaches from coalition logics (e.g., [3, 42, 55]). This paper focuses on the specific frameworks for know-how with a single agent. It is not the intention of the paper to present an exhaustive survey of all the relevant frameworks and justify all the design choices therein.

Table 1 summarizes the main features of the five know-how logics this paper considers.<sup>2</sup> In all cases, we obtain invariance results and prove Hennessy–Milner-style theorems over finite models. Further variations can be treated similarly.

As any notion of bisimulation, the ones presented here provide a tool for comparing the expressivity of different (here: *knowing how*) languages, sharpening in this way our model theoretical understanding of the discussed frameworks. By using them, we show that (i)  $\mathcal{L}_{Khm}$  is strictly more expressive than  $\mathcal{L}_{Kh}$ , (ii)  $\mathcal{L}_{Khw}$  is incomparable to both  $\mathcal{L}_{Khm}$  and  $\mathcal{L}_{Kh}$ , and (iii)  $\mathcal{L}_{Khs}$  is incomparable to  $\mathcal{L}_H$  (with or without the strategy operator). The models for  $\mathcal{L}_{Kh}$ ,  $\mathcal{L}_{Khw}$ ,  $\mathcal{L}_{Khm}$  are quite different from those for  $\mathcal{L}_{Khs}$  and  $\mathcal{L}_H$ ; thus, it is not meaningful to compare expressivity between these two types of *knowing how* logics.

**Outline.** Our work starts in Section 2 by reviewing the definitions of the basic *knowing how* logic and then proposing and studying an appropriate notion of bisimulation. The ideas are extended for dealing with two other closely related *knowing how* logics: the *knowing how* logic with intermediate constraints is investigated in Section 3 and a weaker *knowing how* logic in Section 4. Then, Section 5 moves to a different *knowing how* setting, providing suitable notions of bisimulation for the logic of strategically *knowing how*, and a single-step variant is discussed in Section 6. Also, while we introduce the corresponding notions of bisimulation along the sections, we use them to contrast the expressive power of the discussed frameworks. The introduced bisimulations are then compared with the existing literature (Section 7). Finally, Section 8 concludes with some remarks and future research directions.

<sup>2</sup> Readers are advised to come back to this table after reading the later sections, to understand the features of each logic. Note that  $\mathcal{L}_H$  also includes a strategy operator as in [46].

**§2. Knowing how.**

**2.1. The language  $\mathcal{L}_{Kh}$  and its semantics.** This subsection introduces the most basic *knowing how* setting, presented in [59, 62]. Through the text, let  $\Pi$  be a countable nonempty set of propositional symbols.

**DEFINITION 2.1** (Language  $\mathcal{L}_{Kh}$ ). *Formulas  $\varphi$  of the language  $\mathcal{L}_{Kh}$  are given by the grammar:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \text{Kh}(\varphi, \varphi),$$

with  $p \in \Pi$ . The language  $\mathcal{L}_P$  denotes the propositional fragment of  $\mathcal{L}_{Kh}$ .

The formula  $\text{Kh}(\psi, \varphi)$  expresses that the agent knows how to achieve  $\varphi$  when  $\psi$  is the case. Other Boolean constants and connectives ( $\top, \perp, \vee, \rightarrow, \leftrightarrow$ ) are defined as usual. Additionally, define  $A\varphi := \text{Kh}(\neg\varphi, \perp)$  and  $E\varphi := \neg A\neg\varphi$ . As we will see later,  $A$  is a global universal modality, with  $A\varphi$  stating that  $\varphi$  holds in all the states.

**DEFINITION 2.2** (Relational model). *A relational model for  $\Pi$  is a tuple  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  where*

- $W$  is a nonempty set of possible states (also denoted by  $D(\mathcal{M})$ );
- $\Sigma$  is a nonempty set of basic actions;
- $R : \Sigma \rightarrow \wp(W \times W)$  assigns a binary accessibility relation  $R_a \subseteq W \times W$  to each  $a \in \Sigma$ , (sometimes we will write  $w R_a v$  for  $(w, v) \in R_a$ ); and
- $V : W \rightarrow \wp(\Pi)$  is an atomic valuation.

A pair  $\mathcal{M}, w$  with  $\mathcal{M}$  a relational model and  $w \in D(\mathcal{M})$  is called a *pointed (relational) model*, with  $w$  its *evaluation point*.

Following [59], a relational model is also called an *ability map*. It represents the ability of the agent by indicating not only the actions she can perform at each state, but also the (potentially nondeterministic) outcome of such actions. There can also be terminating states without outgoing relations. Note how the models allow loops, and thus it should not be understood as a temporal model in the traditional sense. Also, since a model describes the agent’s abilities, any *learning* (and, in general, any action through which the abilities of the agent change) should correspond to an operation that transforms the model, typically by changing the relation for atomic actions (thus allowing the existence of new sequences of actions).

To define the semantics for  $\text{Kh}$ , we need a notion of *strong executability* of a plan (a sequence of actions) in a model, which intuitively means that the plan can *always* be executed successfully to the end. Before the formal definition of strong executability, we introduce some basic notions.

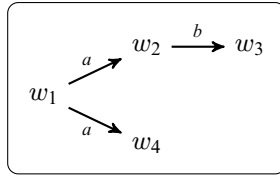
**DEFINITION 2.3.** *Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model for  $\Pi$ ; let  $\sigma \in \Sigma^*$  be a (possibly empty) finite sequence of actions in  $\Sigma$ , with the empty sequence denoted by  $\varepsilon$ . For  $\sigma = \varepsilon$ , define  $R_\varepsilon := \{(w, w) \mid w \in W\}$ ; for  $\sigma = a_1 \dots a_n$ , define*

$$R_{a_1 \dots a_n} := R_{a_1} \circ \dots \circ R_{a_n} \quad \text{where } \circ \text{ is the composition of relations.}$$

Take a state  $w \in W$ . The set  $\{v \in W \mid (w, v) \in R_\sigma\}$ , containing the states  $R_\sigma$ -reachable from  $w$ , is denoted by  $R_\sigma[w]$ .

Note how, if  $\sigma = a \in \Sigma$ , then  $R_\sigma$  is the relation given by the model, as expected.

EXAMPLE 2.4. Consider the following relational model.



According to the previous definition,  $R_\varepsilon = \{(w_1, w_1), (w_2, w_2), (w_3, w_3), (w_4, w_4)\}$ . For  $\sigma_1 = a$ ,  $R_{\sigma_1} = \{(w_1, w_2), (w_1, w_4)\}$ , and for  $\sigma_2 = b$ ,  $R_{\sigma_2} = \{(w_2, w_3)\}$ . Finally,  $R_{ab} = R_{\sigma_1\sigma_2} = R_{\sigma_1} \circ R_{\sigma_2} = \{(w_1, w_3)\}$ .

DEFINITION 2.5 (Executability). Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model for  $\Pi$ . The sequence  $\sigma \in \Sigma^*$  is executable at  $w \in W$  if and only if  $w$  has at least one  $\sigma$ -successor. In other words,  $\sigma$  is executable at  $w \in W$  if and only if  $R_\sigma[w] \neq \emptyset$ .

The notion of strong executability adds one further requirement to Definition 2.5: each time the execution of a strategy starts, it must finish.

DEFINITION 2.6 (Strong Executability). Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model for  $\Pi$ . Given a nonempty sequence  $\sigma = a_1 \dots a_n \in \Sigma^*$  and  $k \leq |\sigma|$ , denote by  $\sigma_k$  the initial segment of  $\sigma$  up to  $a_k$  (i.e.,  $\sigma_k := a_1 \dots a_k$ ). In particular,  $\sigma_0 := \varepsilon$  (for the empty sequence,  $\varepsilon_0 := \varepsilon$ ).

A sequence  $\sigma = a_1 \dots a_n$  is strongly executable (SE) at  $w \in W$  if and only if

$$\text{for any } 0 \leq k < n, \quad v \in R_{\sigma_k}[w] \text{ implies } R_{a_{k+1}}[v] \neq \emptyset.$$

In other words,  $\sigma = a_1 \dots a_n$  is strongly executable at  $w \in W$  if and only if, for any  $0 \leq k < n$ , any state  $R_{\sigma_k}$ -reachable from  $w$  has at least one  $a_{k+1}$ -successor. Note that if  $\sigma$  is strongly executable at  $w$ , then it is executable at  $w$  but the other way around does not hold in general.

EXAMPLE 2.7. Consider the relational model from Example 2.4. Note how  $w_3 \in R_{ab}[w_1]$ , as there is  $w_2$  such that  $w_1 R_a w_2 R_b w_3$ ; then  $R_{ab}[w_1] \neq \emptyset$ , so  $ab$  is executable at  $w_1$ . However,  $ab$  is not strongly executable at  $w_1$ : for its subsequence  $a$  we have  $w_4 \in R_a[w_1]$ , nevertheless  $w_4$  does not have a  $b$ -successor, therefore  $R_b[w_4] = \emptyset$ .

With these tools defined, it is possible to provide the semantic interpretation for the language  $\mathcal{L}_{Kh}$ .

DEFINITION 2.8 (Semantics). Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model, and let  $w \in W$ . The satisfaction relation  $\models$  between the pointed model  $\mathcal{M}, w$  and a formula  $\varphi$  in  $\mathcal{L}_{Kh}$  is inductively defined as follows:

- $\mathcal{M}, w \models p$  iff  $p \in V(w)$
- $\mathcal{M}, w \models \neg\varphi$  iff  $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$  iff  $\mathcal{M}, w \models \varphi$  and  $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models Kh(\psi, \varphi)$  iff there exists a sequence  $\sigma \in \Sigma^*$  such that, for all states  $u \in W$ , if  $\mathcal{M}, u \models \psi$ , then
  - (i)  $\sigma$  is SE at  $u$ , and
  - (ii) for all states  $v \in W, v \in R_\sigma[u]$  implies  $\mathcal{M}, v \models \varphi$ .

A formula  $\varphi$  is satisfiable if and only if there exists a pointed model  $\mathcal{M}, w$  such that  $\mathcal{M}, w \models \varphi$ . As usual,  $\varphi$  is valid if and only if  $\mathcal{M}, w \models \varphi$  for every pointed model

$\mathcal{M}, w$ . Additionally, define the set  $\llbracket \varphi \rrbracket^{\mathcal{M}} := \{w \in W \mid \mathcal{M}, w \models \varphi\}$ , typically called the extension of  $\varphi$  in  $\mathcal{M}$ . The elements of  $\llbracket \varphi \rrbracket^{\mathcal{M}}$  are also referred to as  $\varphi$ -states.

It is worthwhile to notice how, even though the semantic interpretation of  $\text{Kh}$  relies on the existence of a sequence  $\sigma \in \Sigma^*$ , such  $\sigma$  is not a part of the formula. In fact, no formula in  $\mathcal{L}_{\text{Kh}}$  refers to actions in  $\Sigma$ .

For the semantic interpretation of the *knowing how* formulas  $\text{Kh}(\psi, \varphi)$ , the definition establishes that, at state  $w$  of model  $\mathcal{M}$ , the agent knows how to achieve  $\varphi$  given  $\psi$ , written as  $\mathcal{M}, w \models \text{Kh}(\psi, \varphi)$ , if and only if there is a sequence  $\sigma$  that is strongly executable from any  $\psi$ -state, and any of such executions leads to a  $\varphi$ -state. Note how the evaluation point  $w$  does not play any role in the semantic clause, which focuses rather on global model properties. Because of this, a formula of the form  $\text{Kh}(\psi, \varphi)$  is either true in every state (i.e.,  $\llbracket \text{Kh}(\psi, \varphi) \rrbracket^{\mathcal{M}} = D(\mathcal{M})$ ) or else false in every state (i.e.,  $\llbracket \text{Kh}(\psi, \varphi) \rrbracket^{\mathcal{M}} = \emptyset$ ). Thus, this form of *knowing how* relies not on the actual state of affairs, but rather on ‘global’ abilities, just as knowing how to open a safe given that it is closed does not rely on the current state of the safe.

To give the reader a better grasp of the *knowing how* formula  $\text{Kh}(\psi, \varphi)$ , it is useful to delve into the semantic interpretation of the abbreviation  $A\varphi := \text{Kh}(\neg\varphi, \perp)$ . The following proposition states that  $A$  behaves exactly as the global universal modality discussed in [25].

**PROPOSITION 2.9.** *Let  $\mathcal{M}, w$  be a pointed model. Then,  $\mathcal{M}, w \models A\varphi$  if and only if  $\mathcal{M}, u \models \varphi$  holds for all  $u \in D(\mathcal{M})$ .*

*Proof.* Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model; given a sequence  $\sigma \in \Sigma^*$  and a state  $u \in W$ , use  $SE(\sigma, u)$  as a shortcut for ‘ $\sigma$  is strongly executable at  $u$ ’. By unfolding the definition of  $A$ , we get that  $\mathcal{M}, w \models A\varphi$  if and only if

$$\exists \sigma \in \Sigma^*. \forall u \in W. (u \in \llbracket \neg\varphi \rrbracket^{\mathcal{M}} \Rightarrow (SE(\sigma, u) \wedge \forall v \in W. (v \in R_{\sigma}[u] \Rightarrow v \in \llbracket \perp \rrbracket^{\mathcal{M}}))),$$

which reduces to

$$\exists \sigma \in \Sigma^*. \forall u \in W. (u \in \llbracket \neg\varphi \rrbracket^{\mathcal{M}} \Rightarrow (SE(\sigma, u) \wedge \forall v \in W. v \notin R_{\sigma}[u])).$$

Note that the statement  $SE(\sigma, u) \wedge \forall v \in W. v \notin R_{\sigma}[u]$  is contradictory: if  $\sigma \in \Sigma^*$  is  $SE$  at  $u$ , the set  $R_{\sigma}[u]$  cannot be empty. Then, the statement becomes

$$\exists \sigma \in \Sigma^*. \forall u \in W. u \notin \llbracket \neg\varphi \rrbracket^{\mathcal{M}}.$$

Hence, since  $\Sigma^*$  is never empty,

$$\mathcal{M}, w \models A\varphi \quad \text{iff} \quad \text{for all } u \in W, \text{ we have } u \in \llbracket \varphi \rrbracket^{\mathcal{M}}.$$

Thus,  $A$  is the *global* universal modality; as a corollary,  $E\varphi$  is its existential dual. □

For a sound and complete axiomatization of validities in  $\mathcal{L}_{\text{Kh}}$  over the class of relational models, the reader is referred to [59].

In order to finish with the preliminaries, we discuss here two further concepts. First, we introduce the notion of  $\mathcal{L}_{\text{Kh}}$ -equivalence between pointed models.

**DEFINITION 2.10** ( $\mathcal{L}_{\text{Kh}}$ -equivalence). *Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed models. We say  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  are  $\mathcal{L}_{\text{Kh}}$ -equivalent (notation:  $\mathcal{M}, w \equiv_{\mathcal{L}_{\text{Kh}}} \mathcal{M}', w'$ ) if and only if*



they satisfy the same  $\mathcal{L}_{\text{Kh}}$ -formulas, i.e., if and only if, for all  $\varphi$  in  $\mathcal{L}_{\text{Kh}}$ ,

$$\mathcal{M}, w \models \varphi \quad \text{iff} \quad \mathcal{M}', w' \models \varphi.$$

Note that, by Definition 2.8, this is equivalent to checking, for all  $\varphi$  in  $\mathcal{L}_{\text{Kh}}$ , that

$$w \in \llbracket \varphi \rrbracket^{\mathcal{M}} \quad \text{iff} \quad w' \in \llbracket \varphi \rrbracket^{\mathcal{M}'}$$

For the later discussions, we need a notion of definability of a set of states in a model w.r.t. a given formal language:

**DEFINITION 2.11 (Definability).** *Let  $\mathcal{M}$  be a relational model. A set of states  $U \subseteq D(\mathcal{M})$  is  $\mathcal{L}_{\text{Kh}}$ -definable (resp.,  $\mathcal{L}_P$ -definable, or propositionally definable) in  $\mathcal{M}$  if and only if there is  $\varphi \in \mathcal{L}_{\text{Kh}}$  (resp.,  $\varphi \in \mathcal{L}_P$ ) such that  $U = \llbracket \varphi \rrbracket^{\mathcal{M}}$ .*

Since Kh formulas are global (either uniformly true or uniformly false throughout the state space), it is a straightforward exercise to show the following proposition.

**PROPOSITION 2.12.** *Let  $\mathcal{M}$  be a relational model. For all set  $U \subseteq D(\mathcal{M})$ ,  $U$  is  $\mathcal{L}_{\text{Kh}}$ -definable in  $\mathcal{M}$  if and only if it is  $\mathcal{L}_P$ -definable in  $\mathcal{M}$ .*

**2.2. Bisimulation for knowing how.** This section introduces a notion of bisimulation for the *knowing how* language  $\mathcal{L}_{\text{Kh}}$ . We show that this notion is appropriate to prove two classical results: the invariance theorem, and over finite models, a Hennessy-Milner-style theorem. Also, it allows us to compare  $\mathcal{L}_{\text{Kh}}$  with other approaches of knowing how.

When looking for a notion of bisimulation for a given language, one needs to be careful in formulating the conditions: they should be strong enough to guarantee that the language cannot distinguish bisimilar models, but they should also be weak enough to hold between two models that cannot be distinguished by the language. In other words, a bisimulation should be a notion of structural equivalence matching (in appropriate classes of models) the notion of language indistinguishability. In many cases, the definition will look natural after being provided, but the conditions might not have been obvious at all before.

Let us first review the standard notion of bisimulation for multi-modal logic over models with the same set of actions  $\Sigma$  ([12], Section 2.2).

**DEFINITION 2.13 (Bisimulation).** *Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma, R', V' \rangle$  be two relational models. A bisimulation between  $\mathcal{M}$  and  $\mathcal{M}'$  is a nonempty relation  $Z \subseteq W \times W'$  such that  $wZw'$  implies:*

- **Atom:**  $V(w) = V'(w')$ ;
- **Zig:** if  $w R_a v$  for some  $a \in \Sigma$ , then there is a  $v'$  such that  $w' R'_a v'$  and  $vZv'$ ;
- **Zag:** if  $w' R'_a v'$  for some  $a \in \Sigma$ , then there is a  $v$  such that  $w R_a v$  and  $vZv'$ .

We say that  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  are bisimilar ( $\mathcal{M}, w \leftrightarrow \mathcal{M}', w'$ ) when there is a bisimulation  $Z$  between  $\mathcal{M}$  and  $\mathcal{M}'$  such that  $wZw'$ .

It is clear that the standard bisimulation for the basic multi-modal language is not adequate for our purposes. First of all, the standard bisimulation is only defined between models with the same set of actions  $\Sigma$ . However, in our case, the language  $\mathcal{L}_{\text{Kh}}$  does not explicitly talk about the concrete actions in the models. We will define a variant of bisimulation between models with different  $\Sigma$ , which does not match the actions as in the standard bisimulation. As another (obvious) difference, the universal



modality is not definable within the basic multi-modal language, but it is definable within  $\mathcal{L}_{\text{Kh}}$ . Therefore, only matching the connected states is not enough. Still, some of the basic ingredients of the standard bisimulation should appear. From the fact that  $\mathcal{L}_{\text{Kh}}$  involves atomic propositions and Boolean operators, it is obvious that an appropriate bisimulation relation  $Z \subseteq W \times W'$  should only contain pairs of states with matching atomic valuation.

Now, let us discuss the *Zig* and *Zag* conditions (also called *forth* and *back*). They should match the only modal operator of the language, Kh. As we are in a modal setting, one might be tempted to require that, if  $(w, w')$  is in  $Z$ , then these states should have ‘matching’ successors (for an appropriate definition of ‘matching’). However, as it has been emphasized, the actual evaluation point does not play any role in the semantic interpretation of Kh. In fact, as the universal modality is definable in  $\mathcal{L}_{\text{Kh}}$ , every state in  $W$  should have a ‘matching’ state in  $W'$ , and vice-versa, regardless of whether they are accessible from the ‘evaluation point’  $w$ .

But this is clearly not enough: relations between states have not been considered so far, and they are crucial when deciding whether the agent has some strongly executable strategy to achieve a given goal. Further restrictions should be imposed, so the proposed notion of bisimilarity indeed implies  $\mathcal{L}_{\text{Kh}}$ -equivalence.

When looking for these additional conditions, a crucial observation is that the Kh operator does not connect a state with another state (as, e.g., the standard  $\diamond$  and  $\square$  modal operators do); it actually connects a *set of states* (those satisfying the precondition) with another *set of states* (those that can be reached via the strong execution of some given strategy). Conditions similar to what  $\mathcal{L}_{\text{Kh}}$  requires can be found in the literature for bisimulations *over neighbourhood models* (e.g., [5, 7, 27–29, 51, 54]), also related to those for *alternating-time temporal logic* (e.g., [2]) and for *conditional modalities* (e.g., [9]). For this paper’s definition, the following notions will be useful.

**DEFINITION 2.14 (Notation).** *Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model for  $\Pi$ ; take a sequence  $\sigma \in \Sigma^*$  and a set of states  $U \subseteq W$ .*

- Define  $R_\sigma[U]$  as the set of states  $R_\sigma$ -reachable from some element of  $U$ :

$$R_\sigma[U] := \{v \in W \mid u R_\sigma v \text{ for some } u \in U\}.$$

- We write  $U \xrightarrow{\sigma} V$  if and only if

$$(i) \sigma \text{ is strongly executable at every } u \in U \quad \text{and} \quad (ii) V = R_\sigma[U].$$

- We write  $U \rightarrow V$  whenever there is  $\sigma \in \Sigma^*$  such that  $U \xrightarrow{\sigma} V$ .

The relation ‘ $\rightarrow$ ’ between sets of states captures the essence of the modal operator Kh: it relates a set  $U \subseteq W$  with a set  $V \subseteq W$  exactly when there is a sequence in  $\Sigma^*$  that it is strongly executable at every  $u \in U$  and whose execution at some  $u \in U$  leads to some  $v \in V$ . This relation is the crucial notion on the Kh-*Zig* and Kh-*Zag* clauses (those designed to deal with equivalence for Kh-formulas) of the  $\mathcal{L}_{\text{Kh}}$ -bisimulation.

**DEFINITION 2.15.** *Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', R', V' \rangle$  be two relational models; let  $Z \subseteq W \times W'$  be a nonempty relation. Define the set of states in  $W'$  that are reachable via  $Z$  from states in  $U \subseteq W$  as*

$$Z[U] := \{v' \in W' \mid uZv' \text{ for some } u \in U\},$$

and the set of states in  $W$  that are reachable via the converse of  $Z$  from states in  $U' \subseteq W'$  as

$$Z^{-1}[U'] := \{v \in W \mid vZu' \text{ for some } u' \in U'\}.$$

DEFINITION 2.16 ( $\mathcal{L}_{\text{Kh}}$ -bisimulation). Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', R', V' \rangle$  be two relational models. A nonempty relation  $Z \subseteq W \times W'$  is called a  $\mathcal{L}_{\text{Kh}}$ -bisimulation between  $\mathcal{M}$  and  $\mathcal{M}'$  if and only if  $wZw'$  implies all of the following.

- **Atom:**  $V(w) = V'(w')$ .
- **Kh-Zig:** for any propositionally definable set  $U \subseteq W$  in  $\mathcal{M}$ , if  $U \rightarrow V$  for some set  $V \subseteq W$ , then there is a set  $V' \subseteq W'$  such that

$$(i) Z[U] \rightarrow V' \quad \text{and} \quad (ii) V' \subseteq Z[V].$$

- **Kh-Zag:** for any propositionally definable set  $U' \subseteq W'$  in  $\mathcal{M}'$ , if  $U' \rightarrow V'$  for some set  $V' \subseteq W'$ , then there is a set  $V \subseteq W$  such that

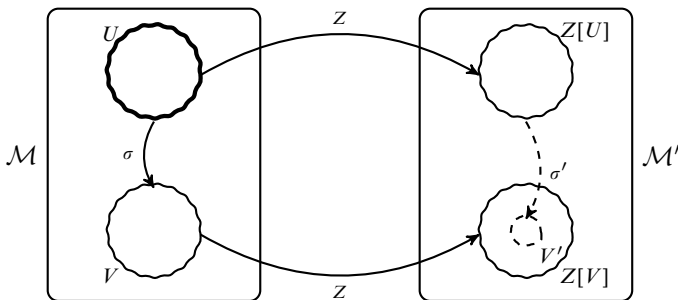
$$(i) Z^{-1}[U'] \rightarrow V \quad \text{and} \quad (ii) V \subseteq Z^{-1}[V'].$$

- **A-Zig:** for every state  $v$  in  $W$  there is a state  $v'$  in  $W'$  such that  $vZv'$ .
- **A-Zag:** for every state  $v'$  in  $W'$  there is a state  $v$  in  $W$  such that  $vZv'$ .

We say  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  are  $\mathcal{L}_{\text{Kh}}$ -bisimilar ( $\mathcal{M}, w \stackrel{\mathcal{L}_{\text{Kh}}}{\leftrightarrow} \mathcal{M}', w'$ ) when there is a  $\mathcal{L}_{\text{Kh}}$ -bisimulation  $Z$  between  $\mathcal{M}$  and  $\mathcal{M}'$  such that  $wZw'$ . We also call the induced  $\stackrel{\mathcal{L}_{\text{Kh}}}{\leftrightarrow}$  relation the  $\mathcal{L}_{\text{Kh}}$ -bisimilarity relation.

The diagram below shows the requirement the Kh-Zig condition imposes (the dashed lines indicate what the condition demands). Let  $U$  be a propositionally definable subset of  $W$ ; if  $U \rightarrow V$  (i.e., if  $V = R_\sigma[U]$  for some  $\sigma \in \Sigma^*$  strongly executable at all  $u \in U$ ), then there should exist a set  $V' \subseteq W'$  such that

- (i)  $Z[U] \rightarrow V'$ , that is,  $V' = R_{\sigma'}[Z[U]]$  (the set  $V'$  contains exactly all those states that are  $\sigma'$ -reachable from some state in  $Z[U]$ ) for some  $\sigma' \in \Sigma'^*$  strongly executable at all  $u' \in Z[U]$  and
- (ii) each state in  $V'$  is the  $Z$ -image of some state in  $V$ .



The Kh-Zag clause works in the other direction: for every propositionally definable set  $U' \subseteq W'$  satisfying  $U' \rightarrow V'$  there should be a set  $V \subseteq W$  such that  $V$  satisfies both (i)  $Z^{-1}[U'] \rightarrow V$ , and (ii)  $V \subseteq Z^{-1}[V']$ . The diagram is analogous to the previous one.

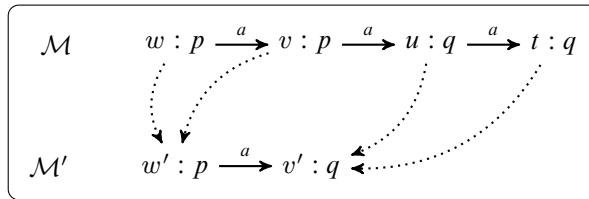
Observe how the Kh-Zig and Kh-Zag clauses of a  $\mathcal{L}_{\text{Kh}}$ -bisimulation work globally with arbitrary propositionally definable subsets of the domain. That they act globally

over sets of states is natural, given the semantic interpretation of  $\mathcal{L}_{\text{Kh}}$ 's only modal operator, Kh. The requirement of *definable* sets is also reasonable, as we are only interested in sets of states the language can distinguish. Finally, requiring *propositionally* definable sets is the same as requiring  $\mathcal{L}_{\text{Kh}}$ -definable ones (Proposition 2.12). The reason for choosing the former in the formulation is that, in this way, a bisimulation is more ‘structural’, defined only in terms of the model (in particular, in terms of valuations).

Because of their definitions, the clauses guarantee that any ‘ $\rightarrow$ ’ transition in one model can be matched by some ‘ $\rightarrow$ ’ transition in the other. This is exactly what is needed, as ‘ $U \rightarrow V$ ’ is essentially the semantic counterpart of the *knowing how* modality  $\text{Kh}(\psi, \varphi)$ : both require the existence of a strongly executable sequence of actions that, when started at any element of one set ( $U$  and  $\llbracket \psi \rrbracket^{\mathcal{M}}$ , respectively), ends up invariably in some element of the other ( $V$  and  $\llbracket \varphi \rrbracket^{\mathcal{M}}$ , respectively).

In the next example we show two bisimilar models, together with the conditions to be checked according to Definition 2.16.

EXAMPLE 2.17. Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', R', V' \rangle$  be the relational models of the diagram below, with nodes specified by both their names and the atoms true at them, and arrows describing the relations  $R_a$  and  $R'_a$ . Take the relation  $Z \subseteq W \times W'$  to be as depicted by the dotted lines.



Note how  $Z$  is a  $\mathcal{L}_{\text{Kh}}$ -bisimulation between  $\mathcal{M}$  and  $\mathcal{M}'$ . First,  $Z$  clearly satisfies the Atom clause, and also A-Zig and A-Zag. Moreover, it satisfies Kh-Zig too: every  $\mathcal{L}_P$ -definable subset  $U \subseteq D(\mathcal{M})$  satisfies the given requirements, as shown in the table below. Therein, we list the sets  $U, V$  and  $V'$  in the model, and also the corresponding bisimulation images  $Z[U]$  and  $Z[V]$ . The second column specifies which  $\mathcal{L}_P$ -formula defines the set  $U$ , whereas the fourth and the sixth columns enumerate the witness sequences for each relation  $U \rightarrow V$  and  $Z[U] \rightarrow V'$ , respectively. Notice that on each row, the set in column  $V'$  is included in the set in column  $Z[V]$ , as requested for the satisfaction of Kh-Zig.

$U$	Def. by	$V$	$U \rightarrow V$ by	$Z[U]$	$Z[U] \rightarrow V'$ by	$V'$	$Z[V]$
$\emptyset$	$\perp$	$\emptyset$	$\varepsilon$	$\emptyset$	$\varepsilon$	$\emptyset$	$\emptyset$
$\{w, v\}$	$p$	$\{w, v\}$	$\varepsilon$	$\{w'\}$	$\varepsilon$	$\{w'\}$	$\{w'\}$
		$\{v, u\}$	$a$	$\{w'\}$	$a$	$\{v'\}$	$\{w', v'\}$
		$\{u, t\}$	$aa$	$\{w'\}$	$a$	$\{v'\}$	$\{v'\}$
$\{u, t\}$	$q$	$\{u, t\}$	$\varepsilon$	$\{v'\}$	$\varepsilon$	$\{v'\}$	$\{v'\}$
$W$	$\top$	$W$	$\varepsilon$	$W'$	$\varepsilon$	$W'$	$W'$

An analogous argument shows that  $Z$  satisfies Kh-Zag too.

Observe also the A-Zig and A-Zag clauses, which simply state that a bisimulation needs to be total for both models (the global universal modality is definable), and which are required to obtain modal equivalence (the proof of Theorem 2.20 will show this). Note how, although A is definable in terms of Kh, the clauses A-Zig and A-Zag do not follow from Kh-Zig and Kh-Zag, as shown by the following example.

EXAMPLE 2.18. Consider the models from Example 2.17. Define  $\mathcal{M}'' = \langle W' \cup \{z\}, \Sigma', R', V'' \rangle$ , where  $V''$  is exactly as  $V'$  except that  $V(z) = \{r\}$ . In other words,  $\mathcal{M}''$  is the result of adding an isolated  $r$ -state into  $\mathcal{M}'$ , namely  $z$ . Let  $Z$  the relation defined as in Example 2.17, now between  $\mathcal{M}$  and  $\mathcal{M}''$ . There is no state  $s \in W$ , s.t.  $sZz$ ; thus, the clauses *Atom*, *Kh-Zig* and *Kh-Zag* still hold. Moreover, *A-Zag* holds too. However, *A-Zig* fails, as no state in  $W$  is sent to  $z$ .

Finally, note how the relation of  $\mathcal{L}_{\text{Kh}}$ -bisimilarity is indeed an equivalence relation among relational models, a fact that is sometimes used implicitly.

PROPOSITION 2.19. The bisimulation relation  $\Leftrightarrow_{\mathcal{L}_{\text{Kh}}}$  is an equivalence relation.

*Proof.* We need to show that  $\Leftrightarrow_{\mathcal{L}_{\text{Kh}}}$  is reflexive, symmetric, and transitive. The first two are straightforward. For transitivity, let  $\mathcal{M}_1 = \langle W_1, R_1, V_1 \rangle$ ,  $\mathcal{M}_2 = \langle W_2, R_2, V_2 \rangle$  and  $\mathcal{M}_3 = \langle W_3, R_3, V_3 \rangle$  be models, with states  $w_1 \in W_1$ ,  $w_2 \in W_2$ ,  $w_3 \in W_3$  satisfying  $\mathcal{M}_1, w_1 \Leftrightarrow_{\mathcal{L}_{\text{Kh}}} \mathcal{M}_2, w_2$  and  $\mathcal{M}_2, w_2 \Leftrightarrow_{\mathcal{L}_{\text{Kh}}} \mathcal{M}_3, w_3$ . Then, there are bisimulations  $Z_1$  and  $Z_2$  such that  $w_1 Z_1 w_2$  and  $w_2 Z_2 w_3$ . We will show that  $Z_2 \circ Z_1$  is a  $\mathcal{L}_{\text{Kh}}$ -bisimulation containing the pair  $(w_1, w_3)$ . The clauses *Atom*, *A-Zig* and *A-Zag* are easy to prove. For *Kh-Zig* we need to show that, for each  $\mathcal{L}_P$ -definable set  $U$  with  $U \rightarrow V$  in  $\mathcal{M}_1$ , there is a set  $V'' \subseteq W_3$  such that  $(Z_2 \circ Z_1)[U] \rightarrow V''$  and  $V'' \subseteq (Z_2 \circ Z_1)[V]$ .

The first step is to show that  $Z_1[U]$  is  $\mathcal{L}_P$ -definable in  $\mathcal{M}_2$ . Let  $\varphi$  be the propositional formula that defines  $U$  in  $\mathcal{M}_1$ ; we claim  $Z_1[U]$  is also definable by  $\varphi$  in  $\mathcal{M}_2$ . First, it is straightforward that all states in  $Z_1[U]$  satisfy  $\varphi$ . Indeed, each state  $u' \in Z_1[U]$  is connected to some state  $u \in U$  by  $Z_1$  so, by *Atom*,  $u$  and  $u'$  should satisfy the same propositional formulas. In particular,  $u'$  satisfies  $\varphi$ . Then, towards a contradiction, suppose there is a state  $u' \in W_2$  that satisfies  $\varphi$  but that is not in  $Z_1[U]$ . By *A-Zag*, there must be a state  $u \in W_1$  such that  $uZ_1u'$ ; by *Atom*,  $u$  satisfies  $\varphi$ . Thus,  $u \in U$  since  $U$  is defined by  $\varphi$  in  $\mathcal{M}_1$ . However, this contradicts the assumption  $u' \notin Z_1[U]$ .

Now we prove *Kh-Zig* for  $Z_2 \circ Z_1$ . From *Kh-Zig* for  $Z_1$ , there exists a set  $V' \subseteq W_2$  such that  $Z_1[U] \rightarrow V'$  in  $\mathcal{M}_2$  and  $V' \subseteq Z_1[V]$ . From *Kh-Zig* for  $Z_2$  and the fact that  $Z_1[U]$  is  $\mathcal{L}_P$ -definable, there exists  $V'' \subseteq W_3$  such that  $Z_2[Z_1[U]] \rightarrow V''$  in  $\mathcal{M}_3$  and  $V'' \subseteq Z_2[V'] \subseteq Z_2[Z_1[V]]$ . Therefore,  $Z_2 \circ Z_1[U] \rightarrow V''$  and  $V'' \subseteq Z_2 \circ Z_1[V]$ .

The case for *Kh-Zag*. Finally, we have  $\mathcal{M}_1, w_1 \Leftrightarrow_{\mathcal{L}_{\text{Kh}}} \mathcal{M}_3, w_3$ . □

In order to simplify some steps in our proofs below, we introduce the following notation:  $\mathcal{M}, w \models \text{Kh}(\psi, \varphi)$  (or, equivalently by Definition 2.8,  $w \in \llbracket \text{Kh}(\psi, \varphi) \rrbracket^{\mathcal{M}}$ ) if and only if

$$\text{there is } \sigma \in \Sigma^* \text{ such that } \llbracket \psi \rrbracket^{\mathcal{M}} \xrightarrow{\sigma} V \text{ and } V = R_{\sigma}[\llbracket \psi \rrbracket^{\mathcal{M}}] \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}.$$

The statement is equivalent to the semantic clause for *Kh* (Definition 2.8):

$$\exists \sigma \in \Sigma^*. \forall u \in W. (u \in \llbracket \psi \rrbracket^{\mathcal{M}} \Rightarrow (SE(\sigma, u) \ \& \ \forall v \in W. (v \in R_{\sigma}[u] \Rightarrow v \in \llbracket \varphi \rrbracket^{\mathcal{M}}))).$$

Both statements are indeed equivalent: from the latter, distribute first the antecedent of the outermost implication, and then the outermost universal quantifier. Using the new notation from Definition 2.14 does the job.

With these tools on the table, it is now time to show that  $\Leftrightarrow_{\mathcal{L}_{\text{Kh}}}$  is indeed an appropriate notion of bisimulation for  $\mathcal{L}_{\text{Kh}}$ . The first result states that  $\mathcal{L}_{\text{Kh}}$ -bisimilarity implies  $\mathcal{L}_{\text{Kh}}$ -equivalence.

**THEOREM 2.20** ( $\mathcal{L}_{\text{Kh}}$  Invariance). *Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed models, with  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', R', V' \rangle$ . Then,*

$$\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Kh}}} \mathcal{M}', w' \text{ implies } \mathcal{M}, w \equiv_{\mathcal{L}_{\text{Kh}}} \mathcal{M}', w'.$$

*Proof.* If  $\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Kh}}} \mathcal{M}', w'$ , then there is a  $\mathcal{L}_{\text{Kh}}$ -bisimulation  $Z \subseteq W \times W'$  such that  $wZw'$ . The proof of  $\mathcal{L}_{\text{Kh}}$ -equivalence is by structural induction on  $\mathcal{L}_{\text{Kh}}$ -formulas. The case for atomic propositions follows immediately from the *Atom* clause, and the cases for the Boolean operators follow from their inductive hypotheses.

The missing piece, the  $\text{Kh}(\psi, \varphi)$  case, uses the auxiliary result  $(\dagger) : Z[\llbracket \psi \rrbracket^{\mathcal{M}}] = \llbracket \psi \rrbracket^{\mathcal{M}'}$ .

- $(\subseteq)$  Take any state  $v' \in Z[\llbracket \psi \rrbracket^{\mathcal{M}}]$ . Then, there is a state  $v \in \llbracket \psi \rrbracket^{\mathcal{M}}$  such that  $vZv'$ . By inductive hypothesis ( $\psi$  is a subformula of  $\text{Kh}(\psi, \varphi)$ ), we have  $v' \in \llbracket \psi \rrbracket^{\mathcal{M}'}$ .
- $(\supseteq)$  Take any state  $v' \in \llbracket \psi \rrbracket^{\mathcal{M}'}$ . By *A-Zag*, there is a state  $v$  with  $vZv'$ ; by inductive hypothesis ( $\psi$  is a subformula of  $\text{Kh}(\psi, \varphi)$ ),  $v \in \llbracket \psi \rrbracket^{\mathcal{M}}$ . Hence,  $v' \in Z[\llbracket \psi \rrbracket^{\mathcal{M}}]$ .

For the actual proof, suppose  $w \in \llbracket \text{Kh}(\psi, \varphi) \rrbracket^{\mathcal{M}}$ . From the semantic interpretation, and using the notation of Definition 2.14, it follows that

$$\text{there is } \sigma \in \Sigma^* \text{ such that } \llbracket \psi \rrbracket^{\mathcal{M}} \xrightarrow{\sigma} V \text{ and } V = R_{\sigma}[\llbracket \psi \rrbracket^{\mathcal{M}}] \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}.$$

The set  $\llbracket \psi \rrbracket^{\mathcal{M}}$  is clearly  $\mathcal{L}_{\text{Kh}}$ -definable, and thus  $\mathcal{L}_P$ -definable (Proposition 2.12). Moreover: there is a sequence  $\sigma \in \Sigma^*$  such that  $\llbracket \psi \rrbracket^{\mathcal{M}} \xrightarrow{\sigma} V$ , so  $\llbracket \psi \rrbracket^{\mathcal{M}} \rightarrow V$ . Then, from clause *Kh-Zig*, there is a set  $V' \subseteq W'$  such that  $(i) Z[\llbracket \psi \rrbracket^{\mathcal{M}}] \rightarrow V'$  (i.e.,  $\llbracket \psi \rrbracket^{\mathcal{M}'} \rightarrow V'$ , given the auxiliary result  $(\dagger)$ ) and  $(ii) V' \subseteq Z[V]$ .

Now, from  $\llbracket \psi \rrbracket^{\mathcal{M}} \rightarrow V$  and  $w \in \llbracket \text{Kh}(\psi, \varphi) \rrbracket^{\mathcal{M}}$  we know that  $V \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}$ . But the inductive hypothesis states that, for proper subformulas of  $\text{Kh}(\psi, \varphi)$ ,  $\mathcal{L}_{\text{Kh}}$ -bisimilarity implies  $\mathcal{L}_{\text{Kh}}$ -equivalence; then, from  $V \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}$  it follows that  $Z[V] \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}'}$ . This, together with  $(ii)$ , yields  $V' \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}'}$ . Thus, we have both  $\llbracket \psi \rrbracket^{\mathcal{M}'} \rightarrow V'$  and  $V' \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}'}$ ; therefore,  $w' \in \llbracket \text{Kh}(\psi, \varphi) \rrbracket^{\mathcal{M}'}$ .

The other direction, from  $w' \in \llbracket \text{Kh}(\psi, \varphi) \rrbracket^{\mathcal{M}'}$  to  $w \in \llbracket \text{Kh}(\psi, \varphi) \rrbracket^{\mathcal{M}}$ , follows a symmetric argument, using *A-Zig* (for the auxiliary result  $Z^{-1}[\llbracket \psi \rrbracket^{\mathcal{M}'}] = \llbracket \psi \rrbracket^{\mathcal{M}}$ ) and *Kh-Zag* (for the actual proof) instead. □

Theorem 2.20 tells us that the semantic condition described by  $\Leftrightarrow_{\mathcal{L}_{\text{Kh}}}$  is strong enough to guarantee  $\mathcal{L}_{\text{Kh}}$ -equivalence. The other direction, showing that  $\Leftrightarrow_{\mathcal{L}_{\text{Kh}}}$  is weak enough to hold between any  $\mathcal{L}_{\text{Kh}}$ -equivalent pointed models, does not hold in general. A typical strategy when dealing with the basic modal language is to focus on a weaker result, showing instead that the desired property holds for *image-finite* models: those in which each state has, for every  $a \in \Sigma$ , only a finite number of  $a$ -successors [12, 13]. Here we will focus rather on models with *finite domain*; this is because the universal modality is definable in our language, and thus a finite domain is required in order to ensure that each state has only finitely many ‘successors’.

**THEOREM 2.21.** *Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed finite-domain models, with  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', R', V' \rangle$ . Then,*

$$\mathcal{M}, w \equiv_{\mathcal{L}_{\text{Kh}}} \mathcal{M}', w' \text{ implies } \mathcal{M}, w \leftrightarrow_{\mathcal{L}_{\text{Kh}}} \mathcal{M}', w'.$$

*Proof.* The strategy is to show that the relation of  $\mathcal{L}_{\text{Kh}}$ -equivalence is already a bisimulation. Thus, define  $Z := \{(v, v') \in (W \times W') \mid \mathcal{M}, v \equiv_{\mathcal{L}_{\text{Kh}}} \mathcal{M}', v'\}$ ; in order to show that  $Z$  satisfies the requirements, take any  $(w, w') \in Z$  (i.e.,  $wZw'$ ).

**Atom:** states  $w$  and  $w'$  agree in all  $\mathcal{L}_{\text{Kh}}$ -formulas, and in particular, in all atoms.

**A-Zig:** Take a state  $v \in W$  and suppose, for the sake of a contradiction, that there is no state  $v' \in W'$  such that  $vZv'$ . Then, from  $Z$ 's definition, for each  $v'_i \in W' = \{v'_1, \dots, v'_n\}$  (recall:  $\mathcal{M}'$  has a finite domain) there is a  $\mathcal{L}_{\text{Kh}}$ -formula  $\theta_i$  such that  $\mathcal{M}, v \models \theta_i$  but  $\mathcal{M}', v'_i \not\models \theta_i$ . Now take  $\theta := \theta_1 \wedge \dots \wedge \theta_n$ . Clearly,  $\mathcal{M}, v \models \theta$ ; however,  $\mathcal{M}', v'_i \not\models \theta$  for each  $v'_i \in W'$ , as each one of them makes ‘its’ conjunct  $\theta_i$  false. Then,  $\mathcal{M}, w \models E\theta$  and  $\mathcal{M}', w' \not\models E\theta$ , contradicting the assumption  $wZw'$ .

**A-Zag:** Analogous to the A-Zig case.

**Kh-Zig:** Take any propositionally definable set  $\llbracket \psi \rrbracket^{\mathcal{M}} \subseteq W$  (thus,  $\psi$  is a propositional formula), and suppose  $\llbracket \psi \rrbracket^{\mathcal{M}} \rightarrow V$  for some set  $V \subseteq W$ . We need to find a set  $V' \subseteq W'$  such that **(i)**  $Z[\llbracket \psi \rrbracket^{\mathcal{M}}] \rightarrow V'$  and **(ii)**  $V' \subseteq Z[V]$ .

First, note that  $\llbracket \psi \rrbracket^{\mathcal{M}'} = Z[\llbracket \psi \rrbracket^{\mathcal{M}}]$ . To show **(i)**, take a state  $u' \in Z[\llbracket \psi \rrbracket^{\mathcal{M}}]$ . Then, there is a state  $u \in \llbracket \psi \rrbracket^{\mathcal{M}}$  with  $uZu'$ . Hence, from  $Z$ 's definition,  $u' \in \llbracket \psi \rrbracket^{\mathcal{M}'}$ . To show **(ii)**, take a state  $u' \in \llbracket \psi \rrbracket^{\mathcal{M}'}$ . From A-Zag there is a state  $u \in W$  with  $uZu'$ . Then, from  $Z$ 's definition,  $u \in \llbracket \psi \rrbracket^{\mathcal{M}}$  and hence  $u' \in Z[\llbracket \psi \rrbracket^{\mathcal{M}}]$ .

Thus, we are actually looking for a set  $V' \subseteq W'$  satisfying both **(i)**  $\llbracket \psi \rrbracket^{\mathcal{M}'} \rightarrow V'$  and **(ii)**  $V' \subseteq Z[V]$ . Now, consider two alternatives.

1. If  $\llbracket \psi \rrbracket^{\mathcal{M}} = \emptyset$ , then  $\llbracket \psi \rrbracket^{\mathcal{M}'} = Z[\llbracket \psi \rrbracket^{\mathcal{M}}] = \emptyset$  and hence taking  $V' = \emptyset$  does the job: clearly, **(i)**  $\emptyset \rightarrow \emptyset$  (due to  $\varepsilon \in \Sigma^*$ ) and **(ii)**  $\emptyset \subseteq Z[V]$ .
2. Otherwise,  $\llbracket \psi \rrbracket^{\mathcal{M}} \neq \emptyset$ . Then, from A-Zag it follows  $Z[\llbracket \psi \rrbracket^{\mathcal{M}}] = \llbracket \psi \rrbracket^{\mathcal{M}'} \neq \emptyset$ . For **(i)**, note that the empty sequence  $\varepsilon \in \Sigma^*$  is such that  $\llbracket \psi \rrbracket^{\mathcal{M}'} \xrightarrow{\varepsilon} \llbracket \psi \rrbracket^{\mathcal{M}'}$ ; thus, taking  $V' = \llbracket \psi \rrbracket^{\mathcal{M}'}$  gives us  $\llbracket \psi \rrbracket^{\mathcal{M}'} \rightarrow V'$ . Still, this only guarantees that there is a  $V'$  satisfying **(i)**, and we need one satisfying both **(i)** and **(ii)**. In order to do this, keep in mind that, due to  $\llbracket \psi \rrbracket^{\mathcal{M}} \neq \emptyset$  (resp.,  $\llbracket \psi \rrbracket^{\mathcal{M}'} \rightarrow V'$ ) and the strong executability requirement,  $V$  (resp.,  $V'$ ) must be nonempty too. To show that there is a set  $V' \subseteq W'$  satisfying both **(i)** and **(ii)**, argue by contradiction. Suppose there is no  $V'$  with the given requirements: each set  $V' \subseteq W'$  satisfying  $\llbracket \psi \rrbracket^{\mathcal{M}'} \rightarrow V'$  is such that  $V' \not\subseteq Z[V]$ . In other words, every set  $V' \subseteq W'$  satisfying  $\llbracket \psi \rrbracket^{\mathcal{M}'} \rightarrow V'$  contains a state  $v'_{v'}$  that is not the  $Z$ -image of some state  $v \in V$ . From  $Z$ 's definition, this means that for each state  $v \in V$  there is a formula  $\theta_{v'}$ , such that  $\mathcal{M}, v \models \theta_{v'}$  but  $\mathcal{M}', v' \not\models \theta_{v'}$ . Since the models are finite, one can define the formulas

$$\theta_{v'} := \bigvee_{v \in V} \theta_{v'} \quad \text{and} \quad \theta := \bigwedge_{\{v' \mid \llbracket \psi \rrbracket^{\mathcal{M}'} \rightarrow v'\}} \theta_{v'}.$$

Note how  $\mathcal{M}, v \models \theta$  holds for all states  $v \in V$ : every such  $v$  satisfies its ‘own’ disjunct  $\theta_{v'}$  in each conjunct  $\theta_{v'}$ . Thus, as  $\llbracket \psi \rrbracket^{\mathcal{M}} \rightarrow V$  and

Kh-formulas are global, we have  $\mathcal{M}, w \models \text{Kh}(\psi, \theta)$ . However, for every set  $V'$  satisfying  $\llbracket \psi \rrbracket^{\mathcal{M}'} \rightarrow V'$  there is a state  $v' \in V'$  such that  $\mathcal{M}', v' \not\models \theta$ : the state  $v'$  that cannot be matched with any state  $v \in V$  falsifies all  $\theta_{V'}^v$ , hence so does the disjunction  $\theta_{V'}$  and therefore also the conjunction  $\theta$ . Thus, again using the fact that Kh-formulas are global,  $\mathcal{M}', w' \not\models \text{Kh}(\psi, \theta)$ , thus contradicting the  $\mathcal{L}_{\text{Kh}}$ -equivalence of  $w$  and  $w'$ .

**Kh-Zag:** Analogous to the Kh-Zig case.

Thus,  $Z$  is a  $\mathcal{L}_{\text{Kh}}$ -bisimulation; hence,  $\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Kh}}} \mathcal{M}', w'$ . □

Note that, without the finite domain condition, Theorem 2.21 does not hold. To see this, consider two pointed models  $\mathcal{M}, w$  and  $\mathcal{M}', w'$ , both over a set of atoms  $\{p_i \mid i \in \mathbb{N}\}$ . Assume that, in  $\mathcal{M}$ , state  $w$  has countably infinite  $a$ -successors  $w_0, w_1, \dots$ , with each atom  $p_k$  true at some  $w_n$  iff  $n \geq k$ ; assume  $\mathcal{M}', w'$  is just as  $\mathcal{M}, w$ , but with  $w'$  having one more successor,  $w_\omega$ , where all the propositional letters are true. It is not hard to see that  $\mathcal{M}, w \equiv_{\mathcal{L}_{\text{Kh}}} \mathcal{M}', w'$  holds (any [recall: finite] sequence starting at  $w'$  can be matched by a sequence starting at  $w$ ), but it is not the case that  $\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Kh}}} \mathcal{M}', w'$  since we can never match  $w_\omega$  in  $\mathcal{M}'$  by A-Zig. To weaken the finiteness condition, some notion of *saturation* is needed, but it is out of the scope of the present paper (cf., e.g., [12]).

It is worthwhile to emphasize that a  $\mathcal{L}_{\text{Kh}}$ -bisimulation works by checking that the discussed models coincide at the abstract level of the ‘ $\rightarrow$ ’ transition (Definition 2.14). This emphasizes the imbalance in the framework (rich models described by weak languages), but also suggests that a better balance can be achieved by using a coarser semantic structure where ‘ $\rightarrow$ ’ is the primitive notion.

**§3. Knowing how with intermediate constraints.** The knowing-how operator Kh has a ‘conditional’ taste:  $\mathcal{M}, w \models \text{Kh}(\psi, \varphi)$  holds if and only if the agent has a ‘method’  $\sigma \in \Sigma^*$  whose output is guaranteed to be  $\varphi$  whenever  $\psi$  holds in the initial situation. Still, the operator is indifferent about the way the given method works: as long as it always takes us from  $\psi$ -states to  $\varphi$ -states, any strongly executable  $\sigma$  will do.

However, in some situations one might be interested not only in the strategy’s final outcome, but also on its intermediate stages. In particular, one might want to guarantee that the strategy is ‘appropriate’ by asking for these intermediate stages to satisfy a certain condition. This is the idea behind the *knowing how* operator with intermediate constraints studied in [38].

**3.1. The language  $\mathcal{L}_{\text{Kh}_m}$  and its semantics.**

DEFINITION 3.22 (Language  $\mathcal{L}_{\text{Kh}_m}$ ). *Formulas  $\varphi$  of the language  $\mathcal{L}_{\text{Kh}_m}$  are given by the grammar:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \text{Kh}_m(\varphi, \varphi, \varphi),$$

with  $p \in \Pi$ . The abbreviation  $A\varphi$  is defined now as  $A\varphi := \text{Kh}_m(\neg\varphi, \top, \perp)$  (with  $E\varphi := \neg A\neg\varphi$ , as before).

In the language  $\mathcal{L}_{\text{Kh}_m}$ , the binary *knowing how* modality Kh from  $\mathcal{L}_{\text{Kh}}$  is replaced by a ternary modality Kh<sub>m</sub>, understood as *knowing how while maintaining*. A formula  $\text{Kh}_m(\psi, \chi, \varphi)$  expresses that, given  $\psi$ , the agent knows how to achieve  $\varphi$  while maintaining  $\chi$ . Formulas of the language are still interpreted in relational models



(Definition 2.2). However, their interpretation requires a refinement of the notion of *strong executability*: now one should take care of what happens at intermediate states.

DEFINITION 3.23. Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model for  $\Pi$ ; let  $\chi$  be a formula in  $\mathcal{L}_{\text{Khm}}$ . A sequence  $\sigma = a_1 \dots a_n \in \Sigma^*$  is strongly  $\chi$ -executable ( $SE\text{-}\chi$ ) at  $w \in W$  if and only if

- (i)  $\sigma$  is strongly executable at  $w$  and
- (ii)  $R_{\sigma_k}[w] \subseteq \llbracket \chi \rrbracket^{\mathcal{M}}$  for every  $0 < k < n$ .

Note how strong  $\chi$ -executability does not require for  $\chi$  to hold at either end of the paths; it is enough for it to hold at the intermediate states. With this definition at hand, we are in a position to introduce the language’s semantic interpretation.

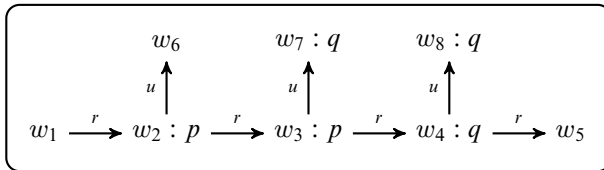
DEFINITION 3.24 (Semantics). Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model, and let  $w \in W$ . The satisfaction relation  $\models$  between the pointed model  $\mathcal{M}, w$  and atoms, negations and conjunctions is as in Definition 2.8. For  $\text{Khm}$ ,

- $\mathcal{M}, w \models \text{Khm}(\psi, \chi, \varphi)$  iff there exists a sequence  $\sigma \in \Sigma^*$  such that, for all states  $u \in W$ , if  $\mathcal{M}, u \models \psi$ , then
- (i)  $\sigma$  is  $SE\text{-}\chi$  at  $u$ , and
  - (ii) for all states  $v \in W, v \in R_{\sigma}[u]$  implies  $\mathcal{M}, v \models \varphi$ .

Note how, under this semantic interpretation, the new definition of the modality  $A$  (as, recall, an abbreviation for  $\text{Khm}(\neg\varphi, \top, \perp)$ ) also makes it the global universal modality: the condition that should be maintained,  $\top$ , is true at every state.

We use an example from [63] to emphasize the differences between the basic *knowing how* operator of the previous section and the one with intermediate constraints discussed in this one.

EXAMPLE 3.25. Consider the following relational model.



Note how, in this model,  $\text{Kh}(p, q)$  holds: if the agent is at a  $p$ -state, she knows how to reach a situation where  $q$  holds. This is because there exists a sequence of actions (namely,  $ru$ ) that, when executed at any  $p$ -state ( $w_2$  or  $w_3$ ), is always completed, and always ending up in a  $q$ -state ( $w_7$  and  $w_8$ , respectively).

However,  $\text{Khm}(p, p, q)$  fails: if the agent is at a  $p$ -state, she does not know how to reach a  $q$ -state by only passing through  $p$ -states. The only strongly executable sequence that ends up in  $q$ -states when executed at  $p$ -ones, namely  $ru$ , has to travel through a  $\neg p$ -state ( $w_4$ ) in the case in which the agent starts at  $w_3$ .

**3.2. Bisimulation for knowing how with intermediate constraints.** Clearly, an appropriate notion of bisimulation for the language  $\mathcal{L}_{\text{Khm}}$  should have some similarities with the notion of bisimulation for  $\mathcal{L}_{\text{Kh}}$  (Definition 2.16). Still, here one should be careful about the states visited by the sequence  $\sigma \in \Sigma^*$  when going from a  $\psi$ -state to one satisfying  $\varphi$ . The definition below refines the previously used concepts by taking into account these intermediate states.

DEFINITION 3.26. Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model for  $\Pi$ ; take  $X \subseteq W$ .

- A sequence  $\sigma = a_1 \dots a_n \in \Sigma^*$  is strongly  $X$ -executable (SE- $X$ ) at state  $w \in W$  if and only if
  - (i)  $\sigma$  is strongly executable at  $w$  and (ii)  $R_{\sigma_k}[w] \subseteq X$  for every  $0 < k < n$ .
- We write  $U \xrightarrow{\sigma, X} V$  if and only if
  - (i)  $\sigma$  is strongly  $X$ -executable at every state  $u \in U$  and (ii)  $V = R_\sigma[U]$ .
- We write  $U \xrightarrow{X} V$  whenever there is a sequence  $\sigma \in \Sigma^*$  such that  $U \xrightarrow{\sigma, X} V$ .

Analogous to the ‘simpler’ knowing how case,  $\xrightarrow{X}$  is the crucial notion on the Khm-Zig and Khm-Zig clauses of a  $\mathcal{L}_{\text{Khm}}$ -bisimulation.

DEFINITION 3.27 ( $\mathcal{L}_{\text{Khm}}$ -bisimulation). Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', R', V' \rangle$  be two relational models. A nonempty relation  $Z \subseteq W \times W'$  is called a  $\mathcal{L}_{\text{Khm}}$ -bisimulation between  $\mathcal{M}$  and  $\mathcal{M}'$  if and only if  $wZw'$  implies all of the following.

- **Atom, A-Zig and A-Zag**, with these clauses as in Definition 2.16.
- **Khm-Zig**: for any propositionally definable set  $U \subseteq W$ , if  $U \xrightarrow{X} V$  for some sets  $X, V \subseteq W$ , then there are sets  $X', V' \subseteq W'$  such that all of the following hold.

$$(i) Z[U] \xrightarrow{X'} V', \quad (ii) X' \subseteq Z[X], \quad (iii) V' \subseteq Z[V].$$

- **Khm-Zag**: for any propositionally definable set  $U' \subseteq W'$ , if  $U' \xrightarrow{X'} V'$  for some sets  $X', V' \subseteq W'$ , then there are sets  $X, V \subseteq W$  such that

$$(i) Z^{-1}[U'] \xrightarrow{X} V, \quad (ii) X \subseteq Z^{-1}[X'], \quad (iii) V \subseteq Z^{-1}[V'].$$

We write  $\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Khm}}} \mathcal{M}', w'$  when there is a  $\mathcal{L}_{\text{Khm}}$ -bisimulation  $Z$  between  $\mathcal{M}$  and  $\mathcal{M}'$  such that  $wZw'$ .

Note the difference between clauses Kh-Zig/Zag and Khm-Zig/Zag: the latter also require for the ‘travelled states’ to be bisimilar. Note also how the *propositional definability* requirement is kept since, just as with  $\mathcal{L}_{\text{Kh}}$ , a set  $U \subseteq D(\mathcal{M})$  is  $\mathcal{L}_{\text{Khm}}$ -definable if and only if it is propositionally definable. On the other hand, we do not need the intermediate condition  $X$  to be propositionally definable in the definition (we only need  $Z[X]$  to be propositional definable in the proof of Theorem 3.29, which can be proved).

The next theorems establish that  $\Leftrightarrow_{\mathcal{L}_{\text{Khm}}}$  is an appropriate notion of bisimulation for  $\mathcal{L}_{\text{Khm}}$  over relational models. For modal equivalence, the notion of  $\mathcal{L}_{\text{Khm}}$ -equivalence ( $\equiv_{\mathcal{L}_{\text{Khm}}}$ ) is defined as in Definition 2.10, but for formulas in  $\mathcal{L}_{\text{Khm}}$ .

First:  $\mathcal{L}_{\text{Khm}}$ -bisimilarity ( $\Leftrightarrow_{\mathcal{L}_{\text{Khm}}}$ ) implies  $\mathcal{L}_{\text{Khm}}$ -equivalence ( $\equiv_{\mathcal{L}_{\text{Khm}}}$ ).

THEOREM 3.28 ( $\mathcal{L}_{\text{Khm}}$  Invariance). Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed models, with  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', R', V' \rangle$ . Then,

$$\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Khm}}} \mathcal{M}', w' \text{ implies } \mathcal{M}, w \equiv_{\mathcal{L}_{\text{Khm}}} \mathcal{M}', w'.$$

*Proof.* From  $\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Khm}}} \mathcal{M}', w'$ , there is a  $\mathcal{L}_{\text{Khm}}$ -bisimulation  $Z \subseteq W \times W'$  such that  $wZw'$ . The proof of  $\mathcal{L}_{\text{Khm}}$ -equivalence is by structural induction on  $\mathcal{L}_{\text{Khm}}$ -formulas, with the case for atomic propositions being immediate from the *Atom* clause, and the cases for Boolean operators following from the inductive hypotheses.

For formulas of the form  $\text{Khm}(\psi, \chi, \varphi)$ , suppose  $w \in \llbracket \text{Khm}(\psi, \chi, \varphi) \rrbracket^{\mathcal{M}}$ . Then, by semantic interpretation, and using the new notation (Definition 3.26),

there is  $\sigma \in \Sigma^*$  such that  $\llbracket \psi \rrbracket^{\mathcal{M}} \xrightarrow{\sigma.X} V$  (for  $X \subseteq \llbracket \chi \rrbracket^{\mathcal{M}}$ ) and  $\mathbf{R}_\sigma[\llbracket \psi \rrbracket^{\mathcal{M}}] \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}$ .

The set  $\llbracket \psi \rrbracket^{\mathcal{M}}$  is clearly  $\mathcal{L}_{\text{Khm}}$ -definable, and thus  $\mathcal{L}_P$ -definable (Proposition 2.12). Moreover: there is a sequence  $\sigma \in \Sigma^*$  such that  $\llbracket \psi \rrbracket^{\mathcal{M}} \xrightarrow{\sigma.X} V$ , so  $\llbracket \psi \rrbracket^{\mathcal{M}} \xrightarrow{X} V$ . Then, from clause Khm-Zig it follows that there are sets  $X', V' \subseteq W'$  such that **(i)**  $Z[\llbracket \psi \rrbracket^{\mathcal{M}}] \xrightarrow{X} V'$ , **(ii)**  $X' \subseteq Z[X]$ , and **(iii)**  $V' \subseteq Z[V]$ .

The following three points will be used to show that  $w' \in \llbracket \text{Khm}(\psi, \chi, \varphi) \rrbracket^{\mathcal{M}'}$ . First: using  $Z[\llbracket \psi \rrbracket^{\mathcal{M}}] = \llbracket \psi \rrbracket^{\mathcal{M}'}$ ,<sup>3</sup> **(i)** yields  $\llbracket \psi \rrbracket^{\mathcal{M}'} \xrightarrow{X} V'$ . Second: using inductive hypothesis on  $\chi$  (for proper subformulas of  $\text{Khm}(\psi, \chi, \varphi)$ ,  $\mathcal{L}_{\text{Khm}}$ -bisimilarity implies  $\mathcal{L}_{\text{Khm}}$ -equivalence), the  $X \subseteq \llbracket \chi \rrbracket^{\mathcal{M}}$  of before implies  $Z[X] \subseteq \llbracket \chi \rrbracket^{\mathcal{M}'}$ ; this, together with **(ii)**, implies  $X' \subseteq \llbracket \chi \rrbracket^{\mathcal{M}'}$ . Third, from the assumed  $w \in \llbracket \text{Khm}(\psi, \chi, \varphi) \rrbracket^{\mathcal{M}}$  and the previous  $\llbracket \psi \rrbracket^{\mathcal{M}} \xrightarrow{X} V$ , the semantic interpretation of Khm implies that  $V \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}$ . By inductive hypothesis on  $\varphi$ , this implies  $Z[V] \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}'}$ ; hence, from **(iii)**, it follows that  $V' \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}'}$ . Thus, summarizing, we have

$$\llbracket \psi \rrbracket^{\mathcal{M}'} \xrightarrow{X} V', \quad X' \subseteq \llbracket \chi \rrbracket^{\mathcal{M}'} \quad \text{and} \quad V' \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}'},$$

which together imply the required  $w' \in \llbracket \text{Khm}(\psi, \chi, \varphi) \rrbracket^{\mathcal{M}'}$ .

The other direction, from  $w' \in \llbracket \text{Khm}(\psi, \chi, \varphi) \rrbracket^{\mathcal{M}'}$  to  $w \in \llbracket \text{Khm}(\psi, \chi, \varphi) \rrbracket^{\mathcal{M}}$ , follows from a symmetric argument, using A-Zig (for the auxiliary result  $Z^{-1}[\llbracket \psi \rrbracket^{\mathcal{M}'}] = \llbracket \psi \rrbracket^{\mathcal{M}}$ ) and Khm-Zag (for the actual proof) instead.  $\square$

Second:  $\mathcal{L}_{\text{Khm}}$ -equivalence ( $\equiv_{\mathcal{L}_{\text{Khm}}}$ ) implies  $\mathcal{L}_{\text{Khm}}$ -bisimilarity ( $\Leftrightarrow_{\mathcal{L}_{\text{Khm}}}$ ) in finite relational models.

**THEOREM 3.29.** *Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed finite-domain models, with  $\mathcal{M} = \langle W, \Sigma, \mathbf{R}, \mathbf{V} \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', \mathbf{R}', \mathbf{V}' \rangle$ . Then,*

$$\mathcal{M}, w \equiv_{\mathcal{L}_{\text{Khm}}} \mathcal{M}', w' \quad \text{implies} \quad \mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Khm}}} \mathcal{M}', w'.$$

*Proof.* As in the proof of Theorem 2.21, we will show that the modal equivalence relation (in this case,  $\equiv_{\mathcal{L}_{\text{Khm}}}$ ) is already a bisimulation. Define  $Z := \{(v, v') \in W \times W' \mid \mathcal{M}, v \equiv_{\mathcal{L}_{\text{Khm}}} \mathcal{M}', v'\}$ , and take any  $(w, w') \in Z$  (i.e.,  $wZw'$ ).

- Atom: states  $w$  and  $w'$  agree in all  $\mathcal{L}_{\text{Khm}}$ -formulas, and in particular, in all atoms.
- A-Zig: As in the proof of Theorem 2.21.
- A-Zag: Ditto.
- Khm-Zig: Take any propositionally definable set  $\llbracket \psi \rrbracket^{\mathcal{M}} \subseteq W$  (so  $\psi$  is a propositional formula), and suppose  $\llbracket \psi \rrbracket^{\mathcal{M}} \xrightarrow{X} V$  for some sets  $X, V \subseteq W$ . We need to find sets  $X', V' \subseteq W'$  satisfying **(i)**  $\llbracket \psi \rrbracket^{\mathcal{M}'} \xrightarrow{X'} V'$ , **(ii)**  $X' \subseteq Z[X]$ , and **(iii)**  $V' \subseteq Z[V]$ .

<sup>3</sup> Its proof, using A-Zag, is as that for the auxiliary result in the proof of Theorem 2.20.

We start by taking the set  $X'$  to be  $Z[X]$ . This gives us (ii) directly, so we only need to show that there is a set  $V' \subseteq W'$  such that

$$(i) \llbracket \psi \rrbracket^{\mathcal{M}'} \xrightarrow{Z[X]} V' \quad \text{and} \quad (ii) V' \subseteq Z[V].$$

This goal is very similar to what the proof of Theorem 2.21 does. Still, in order to use the  $\mathcal{L}_{\text{Khm}}$ -equivalence that  $Z$  entails, we need to show that, in our current case with  $W'$  finite, the set  $Z[X]$  is  $\mathcal{L}_{\text{Khm}}$ -definable. In order to show this, observe that if two states  $u', u'' \in W'$  satisfy exactly the same atoms, then either both are in  $Z[X]$  or else both are outside it. Otherwise, if  $u'$  is the state in  $Z[X]$ , there would be a state  $x \in X$  such that  $xZu'$  holds but  $xZu''$  fails. Since  $Z$  relates  $\mathcal{L}_{\text{Khm}}$ -bisimilar states, this would imply the existence of a formula in  $\mathcal{L}_{\text{Khm}}$  distinguishing  $u''$  from  $x$  ( $xZu''$  fails), and thus distinguishing  $u''$  from  $u'$  ( $xZu'$  holds). But this cannot be the case: we have assumed that  $u'$  and  $u''$  satisfy exactly the same atoms, and thus they satisfy exactly the same Boolean formulas. Moreover, Khm-formulas are global, and thus they cannot distinguish  $u'$  from  $u''$ . After this observation, one can build a propositional formula  $\chi$  characterizing  $Z[X]$  within  $\mathcal{M}'$  (i.e., satisfying  $Z[X] = \llbracket \chi \rrbracket^{\mathcal{M}'}$ ). Indeed, create a partition of  $W'$  such that states in the same partition cell satisfy exactly the same atoms. Since  $W'$  is finite, the number of partition cells is finite, so each one can be distinguished from all the others by means of a finite propositional formula. Then, define  $\chi$  as the (finite) disjunction of the formulas characterizing the partition cells whose union define  $Z[X]$ . Because of the observation, no partition crosses the boundaries between  $Z[X]$  and  $W' \setminus Z[X]$ , and thus  $\chi$  characterizes  $Z[X]$ .

As advanced, the rest of the proof is very similar to the one for Theorem 2.21. In the interesting case ( $\llbracket \psi \rrbracket^{\mathcal{M}} \neq \emptyset$ ), and towards a contradiction, suppose that every set  $V' \subseteq W'$  satisfying  $\llbracket \psi \rrbracket^{\mathcal{M}'} \xrightarrow{Z[X]} V'$  fails to satisfy (iii) for the given set  $V$ . Then, just as in Theorem 2.21, one can find a formula  $\theta \in \mathcal{L}_{\text{Khm}}$  such that  $\mathcal{M}, w \models \text{Khm}(\psi, \chi, \theta)$  and yet  $\mathcal{M}', w' \not\models \text{Khm}(\psi, \chi, \theta)$  (with, remember,  $\chi$  a formula characterizing  $Z[X]$  within  $\mathcal{M}'$ ). Contradiction.

Khm-Zag: Analogous to the Khm-Zig case.

Thus,  $Z$  is a  $\mathcal{L}_{\text{Khm}}$ -bisimulation; therefore,  $\mathcal{M}, w \leftrightarrow_{\mathcal{L}_{\text{Khm}}} \mathcal{M}', w'$ . □

**3.3. Relative expressive power of  $\mathcal{L}_{\text{Kh}}$  and  $\mathcal{L}_{\text{Khm}}$ .** Bisimulations are very useful for investigating the expressive power of a logic. In this subsection, we make use of the tools defined in the previous sections to compare the expressive power of  $\mathcal{L}_{\text{Kh}}$  and  $\mathcal{L}_{\text{Khm}}$ . Notice that in order to compare the expressive power of two frameworks, they should rely on the same class of models. We use the following standard definitions for comparing the expressive power of two languages.

**DEFINITION 3.30.** *Let  $\mathcal{L}_1$  and  $\mathcal{L}_2$  be languages over the same semantic structure.*

- We write  $\mathcal{L}_1 \leq \mathcal{L}_2$ , indicating that  $\mathcal{L}_2$  is at least as expressive as  $\mathcal{L}_1$ , when every formula in  $\mathcal{L}_1$  is semantically equivalent to some formula in  $\mathcal{L}_2$ .
- We write  $\mathcal{L}_1 \not\leq \mathcal{L}_2$ , indicating that  $\mathcal{L}_2$  is not as expressive as  $\mathcal{L}_1$ , when there is a formula in  $\mathcal{L}_1$  that is semantically different from every formula in  $\mathcal{L}_2$ .

- We write  $\mathcal{L}_1 < \mathcal{L}_2$ , indicating that  $\mathcal{L}_2$  is more expressive than  $\mathcal{L}_1$ , when  $\mathcal{L}_1 \leq \mathcal{L}_2$  and  $\mathcal{L}_2 \not\leq \mathcal{L}_1$ .
- We say that  $\mathcal{L}_1$  and  $\mathcal{L}_2$  are incomparable if and only if  $\mathcal{L}_1 \not\leq \mathcal{L}_2$  and  $\mathcal{L}_2 \not\leq \mathcal{L}_1$ .

A typical proof for showing  $\mathcal{L}_1 \leq \mathcal{L}_2$  consists on providing a translation  $\text{tr} : \mathcal{L}_1 \rightarrow \mathcal{L}_2$  that assigns, to any formula  $\varphi$  in  $\mathcal{L}_1$ , a formula  $\text{tr}(\varphi)$  in  $\mathcal{L}_2$  such that, in any given semantic structure for the languages,  $\varphi$  holds if and only if  $\text{tr}(\varphi)$  holds. A typical (simple) proof for showing  $\mathcal{L}_1 \not\leq \mathcal{L}_2$  consists on providing two semantic structures that satisfy exactly the same formulas in  $\mathcal{L}_2$ , and then provide a formula in  $\mathcal{L}_1$  that holds in one structure but fails in the other. In more complicated cases, you may need to find *two classes* of models,  $\{\mathcal{M}_i \mid i \in \mathbb{N}\}$  and  $\{\mathcal{N}_i \mid i \in \mathbb{N}\}$  respectively, such that a formula in  $\mathcal{L}_1$  can distinguish  $\mathcal{M}_i$  and  $\mathcal{N}_i$  for all  $i \in \mathbb{N}$  but any  $\mathcal{L}_2$ -formula cannot distinguish all of them.

PROPOSITION 3.31.  $\mathcal{L}_{\text{Kh}} < \mathcal{L}_{\text{Khm}}$ .

*Proof.* It is easy to see that  $\mathcal{L}_{\text{Kh}} \leq \mathcal{L}_{\text{Khm}}$  since, as stated in [38], we can define the translation  $\text{tr}(\text{Kh}(\psi, \varphi)) := \text{Khm}(\psi, \top, \varphi)$ . For showing  $\mathcal{L}_{\text{Khm}} \not\leq \mathcal{L}_{\text{Kh}}$ , consider the models  $\mathcal{M}$  and  $\mathcal{M}'$  in Example 2.17. As discussed, there is a  $\mathcal{L}_{\text{Kh}}$ -bisimulation between  $\mathcal{M}$  and  $\mathcal{M}'$  containing the pair  $(w, w')$ ; thus, by Theorem 2.20,  $w$  and  $w'$  satisfy the same  $\mathcal{L}_{\text{Kh}}$ -formulas. However, these states can be distinguished by  $\mathcal{L}_{\text{Khm}}$ , as  $\mathcal{M}, w \not\models \text{Khm}(p, \perp, q)$  and  $\mathcal{M}', w' \models \text{Khm}(p, \perp, q)$ .  $\square$

**§4. Weakly knowing how.** The discussed logics of *knowing how* are useful when one is interested in sequences of actions that can be fully executed, and that will end in the desired outcome (the notion of *strong executability*). However, in some situations, this might be a very strong requirement: sometimes, the given sequence might not be executable to its end, and yet it might yield the desired outcome. For example, drinking 10 shots of tequila sounds like a good plan for getting drunk, but the shorter plan consisting of drinking nine shots (or even less) and stop may also achieve the goal.

In [37] the author looked into this alternative by introducing a *knowing how* operator based on *weak conformant* plans. A weak conformant plan for achieving  $\varphi$  from  $\psi$ -states is a finite sequence of actions that, when started on any state satisfying  $\psi$ , it will always end on states satisfying  $\varphi$ , regardless of whether the sequence was executed at its fullest or not.

**4.1. The language  $\mathcal{L}_{\text{Khw}}$  and its semantics.**

DEFINITION 4.32 (Language  $\mathcal{L}_{\text{Khw}}$ ). *Formulas  $\varphi$  of the language  $\mathcal{L}_{\text{Khw}}$  are given by the grammar:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \text{Khw}(\varphi, \varphi),$$

with  $p \in \Pi$ .

Formulas of the form  $\text{Khw}(\psi, \varphi)$  express that, given  $\psi$ , the agent *weakly* knows how to achieve  $\varphi$ . The abbreviation  $A\varphi$  is defined as  $A\varphi := \text{Khw}(\neg\varphi, \perp)$ , as it was done within  $\mathcal{L}_{\text{Kh}}$  (with  $E\varphi := \neg A\neg\varphi$ , as usual).

For  $\text{Khw}$ 's semantic interpretation, a further definition is required.

DEFINITION 4.33. *Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model; take a sequence  $\sigma \in \Sigma^*$  and a state  $w \in W$ . The set  $\text{TerSt}(w, \sigma)$  is the set of states at which the execution of  $\sigma$*

from  $w$  terminates. More precisely,  $\text{TerSt}(w, \varepsilon) := \{w\}$  and

$$\text{TerSt}(w, a_1 \dots a_n) := R_{a_1 \dots a_n}[w] \cup \{v \in W \mid v \in R_{a_1 \dots a_i}[w] \text{ and } R_{a_{i+1}}[v] = \emptyset, \text{ for some } i < n\}.$$

DEFINITION 4.34 (Semantics). Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model; take  $w \in W$ . The satisfaction relation  $\models$  between the pointed model  $\mathcal{M}, w$  and atoms, negations and conjunctions is as in Definition 2.8. For  $\text{Khw}$ ,

$\mathcal{M}, w \models \text{Khw}(\psi, \varphi)$  iff there exists a sequence  $\sigma \in \Sigma^*$  such that, for all states  $u \in W$ , if  $\mathcal{M}, u \models \psi$ , then  $v \in \text{TerSt}(u, \sigma)$  implies  $\mathcal{M}, v \models \varphi$ .

Thus, weakly knowing how to achieve a given  $\varphi$  corresponds to the sketched notion of weak conformant plan: the agent weakly knows how to achieve  $\varphi$  given the condition  $\psi$ , written as  $\text{Khw}(\psi, \varphi)$ , if and only if there is a finite sequence of actions  $\sigma \in \Sigma^*$  that, when started on any state satisfying  $\psi$  (for all  $u \in W$  with  $\mathcal{M}, u \models \psi$ ), it always ends on states satisfying  $\varphi$ , regardless of whether the sequence was executed at its fullest or not ( $v \in \text{TerSt}(u, \sigma)$  implies  $\mathcal{M}, v \models \varphi$ ). Once again, the knowing how modality is global, making  $A$  the universal modality.

EXAMPLE 4.35. In order to illustrate the difference between weakly knowing how and the setting of Section 2, consider again the model in Example 3.25. The agent does not know how to make  $\neg p \wedge \neg q$  true starting from  $p$ -states,  $\neg \text{Kh}(p, \neg p \wedge \neg q)$ , as there is no strongly executable sequence of actions taking her from every state satisfying  $p$  (states  $w_2$  and  $w_3$ ) to only states satisfying  $\neg p \wedge \neg q$  (states  $w_1, w_5$  and  $w_6$ ). Indeed, the only strongly executable sequence going from  $w_3$  to a state satisfying  $\neg p \wedge \neg q$  is  $rr$  (ending at  $w_5$ ), but this sequence does not lead to a  $\neg p \wedge \neg q$ -state from  $w_2$ . However, the agent weakly knows how to reach  $\neg p \wedge \neg q$  given  $p$  (in symbols,  $\text{Khw}(p, \neg p \wedge \neg q)$  holds): when starting from a  $p$ -state, the execution of the sequence  $rrr$  always ends up in a  $(\neg p \wedge \neg q)$ -state, sometimes after a full execution, sometimes after reaching a dead-end.

There are other ways to weaken the executability in the semantics, e.g., [58], where a notion of bisimulation was proposed, inspired by an earlier version of this paper.

4.2. Bisimulation for weakly knowing how. As before, defining a  $\mathcal{L}_{\text{Khw}}$ -bisimulation requires an extra piece of notation.

DEFINITION 4.36. Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  be a relational model; take a sequence  $\sigma \in \Sigma^*$  and a set  $U \subseteq W$ .

- We write  $U \xrightarrow{\sigma}_{\mathcal{W}} V$  if and only if  $V = \bigcup_{u \in U} \text{TerSt}(u, \sigma)$ .
- We write  $U \rightarrow_{\mathcal{W}} V$  if and only if there is a sequence  $\sigma \in \Sigma^*$  such that  $U \xrightarrow{\sigma}_{\mathcal{W}} V$ .

Below we introduce the notion of bisimulation for  $\mathcal{L}_{\text{Khw}}$ .

DEFINITION 4.37 ( $\mathcal{L}_{\text{Khw}}$ -bisimulation). Let  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', R', V' \rangle$  be two relational models. A nonempty relation  $Z \subseteq W \times W'$  is called a  $\mathcal{L}_{\text{Khw}}$ -bisimulation between  $\mathcal{M}$  and  $\mathcal{M}'$  if and only if  $wZw'$  implies all of the following.

- **Atom, A-Zig and A-Zag**, with these clauses as in Definition 2.16.
- **Khw-Zig**: for any propositionally definable set  $U \subseteq W$ , if  $U \rightarrow_{\mathcal{W}} V$  for some set  $V \subseteq W$ , then there is a set  $V' \subseteq W'$  such that

$$(i) Z[U] \rightarrow_{\mathcal{W}'} V' \quad \text{and} \quad (ii) V' \subseteq Z[V].$$

- **Kh<sub>m</sub>-Zag:** for any propositionally definable set  $U' \subseteq W'$ , if  $U' \rightarrow_W V'$  for some set  $V' \subseteq W'$ , then there is a set  $V \subseteq W$  such that

$$(i) Z^{-1}[U'] \rightarrow_W V \quad \text{and} \quad (ii) V \subseteq Z^{-1}[V'].$$

We write  $\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Khw}}} \mathcal{M}', w'$  when there is a  $\mathcal{L}_{\text{Khw}}$ -bisimulation  $Z$  between  $\mathcal{M}$  and  $\mathcal{M}'$  such that  $wZw'$ .

By reasoning as in the proofs of Theorems 2.20 and 3.28, it can be proved that  $\mathcal{L}_{\text{Khw}}$ -bisimilarity ( $\Leftrightarrow_{\mathcal{L}_{\text{Khw}}}$ ) implies  $\mathcal{L}_{\text{Khw}}$ -equivalence ( $\equiv_{\mathcal{L}_{\text{Khw}}}$ ), with the latter defined as in Definition 2.10 but for formulas in  $\mathcal{L}_{\text{Khw}}$ .

**THEOREM 4.38** ( $\mathcal{L}_{\text{Khw}}$  Invariance). *Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed models, with  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', R', V' \rangle$ . Then,*

$$\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Khw}}} \mathcal{M}', w' \quad \text{implies} \quad \mathcal{M}, w \equiv_{\mathcal{L}_{\text{Khw}}} \mathcal{M}', w'.$$

Then, just as for  $\mathcal{L}_{\text{Kh}}$  and  $\mathcal{L}_{\text{Khm}}$  (Theorems 2.21 and 3.29, respectively), in finite-domain relational models,  $\mathcal{L}_{\text{Khw}}$ -equivalence implies  $\mathcal{L}_{\text{Khw}}$ -bisimilarity.

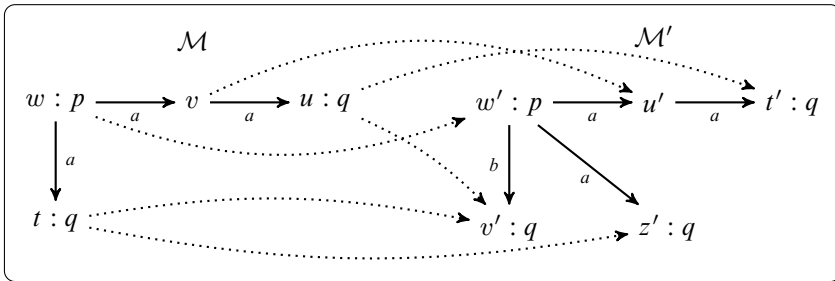
**THEOREM 4.39.** *Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed finite-domain models, with  $\mathcal{M} = \langle W, \Sigma, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', R', V' \rangle$ . Then,*

$$\mathcal{M}, w \equiv_{\mathcal{L}_{\text{Khw}}} \mathcal{M}', w' \quad \text{implies} \quad \mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{\text{Khw}}} \mathcal{M}', w'.$$

**4.3. Expressive power of  $\mathcal{L}_{\text{Khw}}$ .** In this section we compare the expressive power of  $\mathcal{L}_{\text{Khw}}$  with respect to  $\mathcal{L}_{\text{Kh}}$  and  $\mathcal{L}_{\text{Khm}}$ .

**PROPOSITION 4.40.**  $\mathcal{L}_{\text{Kh}} \not\leq \mathcal{L}_{\text{Khw}}$ . Moreover, together with Proposition 3.31, we get  $\mathcal{L}_{\text{Khm}} \not\leq \mathcal{L}_{\text{Khw}}$ .

*Proof.* Consider the following models:



Dotted lines define a  $\mathcal{L}_{\text{Khw}}$ -bisimulation. The table below shows that the models satisfy the requirements for Khw-Zig from Definition 4.37 (Khw-Zag can be similarly checked). Note that we only consider the non-trivial cases.

$U$	Def. by	$V$	$U \rightarrow_W V$ by	$Z[U]$	$V'$	$Z[V]$
$\{w\}$	$p$	$\{v, t\}$	$a$	$\{w'\}$	$\{u', z'\}$	$\{v', u', z'\}$
		$\{t, u\}$	$aa$	$\{w'\}$	$\{t', z'\}$	$\{t', v', z'\}$
$\{v\}$	$\neg p \wedge \neg q$	$\{u\}$	$a$	$\{u'\}$	$\{t'\}$	$\{v', t'\}$

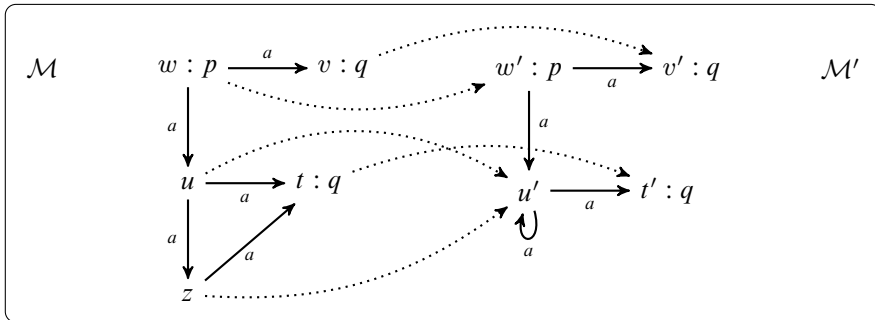


Since  $\mathcal{M}, w$  and  $\mathcal{M}', w$  are  $\mathcal{L}_{K_{hw}}$ -bisimilar, they satisfy the same  $\mathcal{L}_{K_{hw}}$ -formulas (Theorem 4.38). However, they can be distinguished by  $\mathcal{L}_{K_h}$ , as  $\mathcal{M}, w \not\models Kh(p, q)$  but  $\mathcal{M}', w' \models Kh(p, q)$ .  $\square$

One might be tempted to think, then, that both  $\mathcal{L}_{K_h}$  and  $\mathcal{L}_{K_{hm}}$  are strictly more expressive than  $\mathcal{L}_{K_{hw}}$ , but actually this is not the case.

PROPOSITION 4.41.  $\mathcal{L}_{K_{hw}} \not\leq \mathcal{L}_{K_h}$  and  $\mathcal{L}_{K_{hw}} \not\leq \mathcal{L}_{K_{hm}}$ .

*Proof.* Consider the following models:



Dotted lines define both a  $\mathcal{L}_{K_h}$ -bisimulation and a  $\mathcal{L}_{K_{hm}}$ -bisimulation (the only potentially meaningful differences between the models, the  $a$ -transition from  $u$  to  $z$  in  $\mathcal{M}$  and the  $a$ -transition from  $u'$  to itself in  $\mathcal{M}'$ , can mimic each other). Thus, the models cannot be distinguished by  $\mathcal{L}_{K_h}$ , and neither by  $\mathcal{L}_{K_{hm}}$ . However,  $\mathcal{L}_{K_{hw}}$  can tell them apart, as  $\mathcal{M}, w \models Khw(p, q)$  but  $\mathcal{M}', w' \not\models Khw(p, q)$ .  $\square$

As a consequence, we get the following result:

COROLLARY 4.42.  $\mathcal{L}_{K_{hw}}$  is incomparable with both  $\mathcal{L}_{K_h}$  and  $\mathcal{L}_{K_{hm}}$ .

**§5. Strategically knowing how.** The previous sections have introduced bisimulations for different variants of *knowing how* operators, together with their corresponding theorems matching bisimulation with logical equivalence. In all cases, we deal with global modalities, whose truth-value does not depend on the evaluation point. Moreover, these *knowing how* frameworks do not make use of the classical epistemic notion of *knowing that*. As a consequence of the latter, an agent cannot rely on her knowledge about the current situation to decide which action should be executed to achieve the desired outcome. In other words, she cannot have *knowledge-based* plans.

In [21] a logic that puts together the two aspects mentioned above was presented.<sup>4</sup> It uses two epistemic notions, *knowing how* and *knowing that*, with their interaction used to represent conditional plans. To achieve the interaction, the *knowing how* operator becomes local, since now it needs to make use of the epistemic information the agent has in a particular state.

<sup>4</sup> Similar ideas have been presented in [42, 46].

**5.1. The language  $\mathcal{L}_{\text{Khs}}$  and its semantics.**

DEFINITION 5.43 (Language). *The language  $\mathcal{L}_{\text{Khs}}$  has two modalities, K for knowing that, and Khs for knowing how. Formally, formulas  $\varphi$  of  $\mathcal{L}_{\text{Khs}}$  are given by the grammar:*

$$\varphi := \perp \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \text{K}\varphi \mid \text{Khs}(\varphi),$$

with  $p \in \Pi$ .

Note again how no formula in  $\mathcal{L}_{\text{Khs}}$  refers to actions in  $\Sigma$ . More importantly,  $\mathcal{L}_{\text{Khs}}$  has two epistemic operators, and each of them needs to be interpreted on a different dimension. For the Khs modality, the model must have one dimension representing ‘strategies’ (the *ability map* of before), whereas for K we need the classical indistinguishability relation used in classical epistemic logic. This produces richer models, defined formally in the following way.

DEFINITION 5.44 (Bi-dimensional models). *A bi-dimensional model  $\mathcal{M}$  for  $\Pi$  is a tuple  $\langle W, \Sigma, \sim, R, V \rangle$  where  $\langle W, \Sigma, R, V \rangle$  is a relational model for  $\Pi$ , and  $\sim \subseteq W \times W$  is an equivalence relation.*

Note how the relations  $\sim$  and  $R_a$  are independent from each other, thus giving us the most general framework regarding the interaction between the two.

REMARK 5.45. A bi-dimensional model captures, at any state, both the uncertainty the agent might have as well as the actions available to her. In this sense, it is very similar to the models for epistemic temporal logic [50, 55]. By having both the epistemic relation and action transitions in the model, one can represent the way actions change the agent’s knowledge (e.g., she considers more than one epistemic possibility, but in the state that results from executing action  $a$ , only one epistemic possibility remains). However, the Kh-setting does not provide information about the way the agent’s epistemic uncertainty evolves through the execution of a sequence of actions. One way to establish this connection is through the ‘dynamic view’, where the epistemic uncertainty after an action is computed from the epistemic uncertainty before the action. Typically this consists of making two assumptions: perfect recall and no miracles [40, 63].

Let us start by defining the transitions between indistinguishable epistemic states.

DEFINITION 5.46 (Transitions). *Let  $\mathcal{M} = \langle W, \Sigma, \sim, R, V \rangle$  be a bi-dimensional relational model for  $\Pi$ ; take a state  $w \in W$ .*

- *The equivalence class of  $w$  with respect to  $\sim$  is given by  $[w] := \{v \in W \mid w \sim v\}$ . Then,  $W/\sim$  is the collection of all the equivalence classes on  $W$  with respect to  $\sim$ .*
- *We use  $[w]R_a[v]$  to indicate that there are  $w' \in [w]$  and  $v' \in [v]$  such that  $w'R_a v'$ .*

In the next definition, we make precise the notion of strategy.

DEFINITION 5.47 (Strategy). *Let  $\mathcal{M} = \langle W, \Sigma, \sim, R, V \rangle$  be a bi-dimensional relational model for  $\Pi$ . A (uniformly executable) strategy over  $\mathcal{M}$  is a partial function  $\sigma : W/\sim \rightarrow \Sigma$  such that, for every equivalence class  $[w] \in W/\sim$ , the action symbol  $\sigma([w])$  is executable at every state  $w' \in [w]$ .*

A strategy assigns an action not to a state, but rather to a  $\sim$ -equivalence class; this we call *uniformity*. Equally important is the fact that the assigned action symbol should

be executable in every state of the  $\sim$ -equivalence class; without it, the knowledge-how may be trivialized. Note how the empty function is also a strategy: the empty strategy. We use  $\text{dom}(\sigma)$  to denote the domain of  $\sigma$  (i.e., the set of equivalence classes in  $W/\sim$  on which  $\sigma$  is defined).

**DEFINITION 5.48** (Executions over equivalence classes). *Let  $\sigma$  be a strategy over the bi-dimensional model  $\mathcal{M} = \langle W, \Sigma, \sim, \mathbf{R}, \mathbf{V} \rangle$ .*

- *A possible execution of  $\sigma$  is a possibly infinite sequence of equivalence classes*

$$\delta = [w_0][w_1] \dots$$

*such that  $[w_i] \mathbf{R}_{\sigma([w_i])} [w_{i+1}]$  for all  $0 \leq i < |\delta|$ . In particular,  $[w]$  is a possible execution of  $\sigma$  when  $[w] \notin \text{dom}(\sigma)$ .*

- *If a possible execution of a strategy is a finite sequence  $[w_0] \dots [w_n]$ , the last set  $[w_n]$  is called the leaf-node, and each set  $[w_i]$  (for  $0 \leq i < n$ ) is called inner-node, both with respect to this possible execution (note: no equivalence class can be both the leaf-node and an inner-node of a given possible execution). If the execution is infinite, every  $[w_i]$  is an inner-node (but none is a leaf-node).*
- *A possible execution of  $\sigma$  is complete if it is infinite or its leaf-node is not in  $\text{dom}(\sigma)$ .*
- *We use  $\text{CELeaf}(\sigma, w)$  to denote the set of all leaf-nodes of all the  $\sigma$ 's complete executions starting from  $[w]$ .*<sup>5</sup>

**DEFINITION 5.49** (Semantics). *Let  $\mathcal{M} = \langle W, \Sigma, \sim, \mathbf{R}, \mathbf{V} \rangle$  be a bi-dimensional relational model; take a state  $w \in W$ . The satisfaction relation  $\models$  between the pointed model  $\mathcal{M}, w$  and a formula  $\varphi$  in  $\mathcal{L}_{\text{Khs}}$  is inductively defined as before for atoms, negation and conjunction; for the new modalities,*

$$\begin{aligned} \mathcal{M}, w \models \text{K}\varphi & \quad \text{iff} \quad \text{for all } w' : w \sim w' \text{ implies } \mathcal{M}, w' \models \varphi \\ \mathcal{M}, w \models \text{Khs}(\varphi) & \quad \text{iff} \quad \text{there exists a strategy } \sigma \text{ such that} \\ & \quad \text{(i)} \quad [v] \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}} \text{ for all classes } [v] \in \text{CELeaf}(\sigma, w) \\ & \quad \text{(ii)} \quad \text{all its complete executions} \\ & \quad \quad \text{starting from } [w] \text{ are finite.} \end{aligned}$$

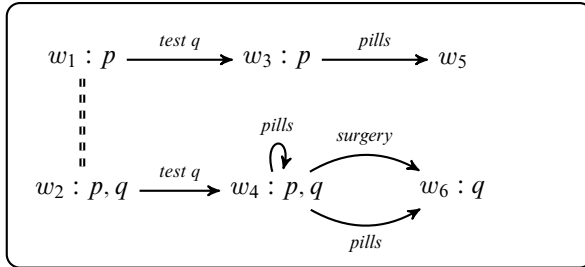
As we already mentioned, with this new logic we are capturing two epistemic properties, unlike the logics presented in previous sections. Notice that although the global universal modality definable in the previous frameworks can be viewed as a restricted form of *knowing that* operator, it can only express global background knowledge to some extent. Moreover, bi-dimensional models contain both epistemic uncertainty and action transitions; thus, one can talk about changes of knowledge after actions, which is essential to facilitate strategies based on local knowledge, compared to the quite restricted linear plans used by [59, 62].

Finally, recall that the global operator  $\text{Kh}(\psi, \varphi)$  of Section 2 expresses that the agent knows how to achieve  $\varphi$  given the condition  $\psi$ . From the perspective of this section's approach, the condition  $\psi$  in the basic *knowing how* setting defines an initial uncertainty set, and  $\text{Kh}(\psi, \varphi)$  can be read rather as '*the agent knows how to achieve  $\varphi$  given that (s)he only knows  $\psi$* '. However, one might consider this a very weak notion

<sup>5</sup> Note how, because of the potential nondeterminism of each relation  $\mathbf{R}_a$ , there might be more than one complete execution of  $\sigma$  starting from a given  $[w]$ .

of uncertainty. The *strategically knowing how* of this section deals with this issue by combining local modalities for different kinds of knowledge.

EXAMPLE 5.50. Consider the bi-dimensional model below [22], example 1.1) (double dashed lines represent  $\sim$ ; reflexive  $\sim$ -edges omitted).



The model represents a scenario in which a doctor needs a plan to treat a patient and cure his pain (represented by the propositional symbol  $p$ ), under the uncertainty about some possible allergy ( $q$ ). On the one hand, if there is no allergy (i.e.,  $q$  fails) then taking some pills will cure the pain, with a surgery not being an option. On the other hand, in presence of the allergy (i.e.,  $q$  holds), the pills may cure the pain or have no effect at all, while the surgery can cure the pain for sure. Notice that there is an action to test whether  $q$  is the case. The indistinguishability between  $w_1$  and  $w_2$  represents the initial uncertainty about the allergy. According to the model, in order to cure the pain (reaching  $\neg p$ ), it makes sense to take the surgery if the test of whether  $q$  is positive, and take the pills otherwise. In this case, the doctor knows how to cure the pain, denoted  $\text{Khs}(\neg p)$ , thanks to the aforementioned strategic plan.

REMARK 5.51. Readers who are familiar with Alternating-time Temporal Logic may see similarities between  $\text{Khs}\varphi$  and  $\text{K}\langle\langle i \rangle\rangle F\varphi$ . However, the frameworks are quite different: here, we need to put the quantifier on strategies outside the knowledge operator (see [22, 31] for discussions). Moreover, we require strict termination of the strategies, so just passing a  $\varphi$ -state is not enough: you have to stop exactly at  $\varphi$ -states.

5.2. **Bisimulation for strategically knowing how.** First, we introduce some further notation.

DEFINITION 5.52. Let  $\mathcal{M} = \langle W, \Sigma, \sim, \mathbf{R}, \mathbf{V} \rangle$  be a model for  $\Pi$ ; take a state  $w \in W$ , a set  $U \subseteq W$ , and a (uniformly executable) strategy  $\sigma$ . We write  $w \xrightarrow{\sigma} U$  when:

1. all complete executions of  $\sigma$  starting from  $[w]$  are finite and
2.  $U = \bigcup_{[v] \in \text{CELeaf}(\sigma, w)} [v]$ .

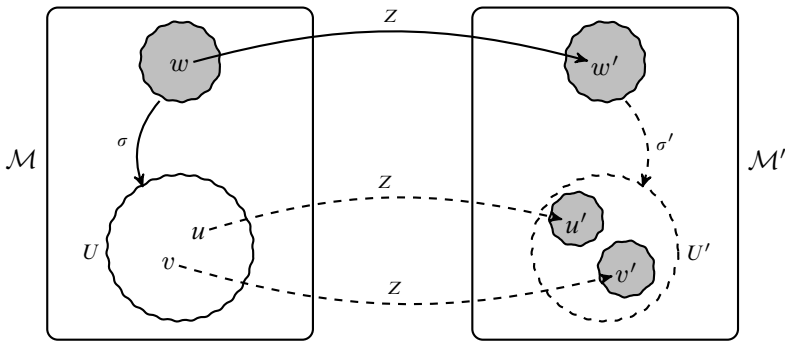
We say that  $w$  strategically reaches  $U$  (denoted  $w \rightsquigarrow U$ ) if there exists a strategy  $\sigma$  such that  $w \xrightarrow{\sigma} U$ .

Then, we introduce a notion of bisimulation for  $\mathcal{L}_{\text{Khs}}$  over bi-dimensional models.

DEFINITION 5.53 ( $\mathcal{L}_{\text{Khs}}$ -bisimulation). Let  $\mathcal{M} = \langle W, \Sigma, \sim, \mathbf{R}, \mathbf{V} \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', \sim', \mathbf{R}', \mathbf{V}' \rangle$  be two bi-dimensional relational models for  $\Pi$ . A nonempty relation  $Z \subseteq W \times W'$  is called a  $\mathcal{L}_{\text{Khs}}$ -bisimulation between  $\mathcal{M}$  and  $\mathcal{M}'$  if and only if  $wZw'$  implies all of the following.

- **Atom:**  $V(w) = V'(w')$ .
- **Zig:** if  $w \sim v$  then there is some state  $v'$  such that  $w' \sim v'$  and  $vZv'$ .
- **Zag:** if  $w' \sim v'$  then there is some state  $v$  such that  $w \sim v$  and  $vZv'$ .
- **Khs-Zig:** if  $w \rightsquigarrow U$  for some set  $U \subseteq W$ , then  $w' \rightsquigarrow U'$  for some set  $U' \subseteq W'$  satisfying  $U' \subseteq Z[U]$ .
- **Khs-Zag:** if  $w' \rightsquigarrow U'$  for some set  $U' \subseteq W'$ , then  $w \rightsquigarrow U$  for some set  $U \subseteq W$  satisfying  $U \subseteq Z^{-1}[U']$ .

The Zig and Zag conditions are inherited from the standard modal bisimulation [12], Section 2.2). The diagram below describes the way the Khs-Zig condition works. Gray circles represent the equivalence classes  $[w]$ ,  $[w']$ ,  $[v']$  and  $[u']$ , for the relations  $\sim$  and  $\sim'$ , respectively. Given  $w \rightsquigarrow U$  (through the strategy  $\sigma$ ), then  $w' \rightsquigarrow U'$  should be the case (for some  $U'$  and through some  $\sigma'$ ). Moreover, each state in  $U'$  (which, recall, might be the union of different equivalence classes due to the potential nondeterminism of the relations), must be the Z-image of some state in  $U$  (so  $vZv'$  and  $uZu'$ ). Clause Khs-Zag works analogously.



Notice that an  $\mathcal{L}_{Khs}$ -bisimulation takes care of two aspects. On the one hand,  $\mathcal{L}_{Khs}$  includes the standard *knowing that* operator  $K$ , so a  $\mathcal{L}_{Khs}$ -bisimulation matches pairs of states when they coincide in the structural information for their respective indistinguishability relations. The usual Zig and Zag conditions take care of this. On the other hand, only states coinciding in their executability should be matched. Unlike the bisimulations introduced before, the *strategic reachability* condition establishes a constraint not over a set of propositionally definable states, but on the states which are epistemically indistinguishable. This way, we are able to describe the proper interaction between the two kinds of knowledge the language considers.

The next theorem establishes that  $\Leftrightarrow_{\mathcal{L}_{Khs}}$  is an appropriate notion of bisimulation for  $\mathcal{L}_{Khs}$ . As usual, the notion of  $\mathcal{L}_{Khs}$ -equivalence,  $\equiv_{\mathcal{L}_{Khs}}$ , is defined as in Definition 2.10, but for formulas in  $\mathcal{L}_{Khs}$ .

**THEOREM 5.54** ( $\mathcal{L}_{Khs}$  Invariance). *Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed bi-dimensional models, with  $\mathcal{M} = \langle W, \Sigma, \sim, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', \sim', R', V' \rangle$ . Then,*

$$\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{Khs}} \mathcal{M}', w' \text{ implies } \mathcal{M}, w \equiv_{\mathcal{L}_{Khs}} \mathcal{M}', w'.$$

*Proof.* Suppose  $\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{Khs}} \mathcal{M}', w'$ ; then there is a  $\mathcal{L}_{Khs}$ -bisimulation  $Z \subseteq W \times W'$  such that  $wZw'$ . The proof is by structural induction on  $\mathcal{L}_{Khs}$ -formulas. Atoms and Boolean cases are straightforward; for the epistemic cases, proceed as follows.

**Case  $K\varphi$ :** Suppose  $\mathcal{M}, w \models K\varphi$ , then for all states  $v \in W$ , if  $w \sim v$  then  $\mathcal{M}, v \models \varphi$ . Now take any state  $v \in W$  such that  $w \sim v'$ ; from **Zig**, there is a state  $v' \in W'$  such that  $w' \sim' v'$  and  $vZv'$ . From  $vZv'$  and inductive hypothesis,  $\mathcal{M}', v' \models \varphi$ . Since  $v'$  is an arbitrary  $\sim'$ -successor of  $w'$ , it follows that  $\mathcal{M}', w' \models K\varphi$ . The reasoning for other direction is analogous, using **Zag** instead.

**Case  $Khs(\varphi)$ :** Suppose  $\mathcal{M}, w \models Khs(\varphi)$ . Then, there exists a strategy  $\sigma$  executable at  $w$  such that

1.  $[t] \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}$  for all classes  $[t] \in \text{CELeaf}(\sigma, w)$  and
2. all its complete executions from  $[w]$  are finite.

Take the set  $U$  such that  $w \xrightarrow{\sigma} U$ . By **Khs-Zig**, there is a strategy  $\sigma'$  such that  $w' \xrightarrow{\sigma'} U'$  for some set  $U' \subseteq W'$  satisfying  $U' \subseteq Z[U]$ . Aiming for contradiction, suppose  $\mathcal{M}', v' \models \neg\varphi$  for some state  $v' \in U'$ ; then, by inductive hypothesis,  $\mathcal{M}, v \models \neg\varphi$  for some state  $v \in U$ . This contradicts the fact  $[v] \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}$  (recall: from  $\xrightarrow{\sigma}$ 's definition, each state  $v \in U$  is in  $[t]$  for some  $[t] \in \text{CELeaf}(\sigma, w)$ ). Therefore,  $\mathcal{M}', v' \models \varphi$ , for all states  $v' \in U'$ . Then,  $\mathcal{M}', w' \models Khs(\varphi)$ . The other direction follows a symmetric argument, using **Khs-Zag** instead. □

We consider now the converse of Theorem 5.54; in finite models,  $\mathcal{L}_{Khs}$ -equivalence ( $\equiv_{\mathcal{L}_{Khs}}$ ) implies  $\mathcal{L}_{Khs}$ -bisimilarity ( $\Leftrightarrow_{\mathcal{L}_{Khs}}$ ).

**THEOREM 5.55.** *Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed finite-domain bi-dimensional models, with  $\mathcal{M} = \langle W, \Sigma, \sim, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', \sim', R', V' \rangle$ . Then,*

$$\mathcal{M}, w \equiv_{\mathcal{L}_{Khs}} \mathcal{M}', w' \text{ implies } \mathcal{M}, w \Leftrightarrow_{\mathcal{L}_{Khs}} \mathcal{M}', w'.$$

*Proof.* Define the relation  $Z := \{(w, w') \in (W \times W') \mid \mathcal{M}, w \equiv_{\mathcal{L}_{Khs}} \mathcal{M}', w'\}$ . We will show that  $Z$  is a bisimulation, so take any  $(w, w') \in Z$  (i.e.,  $wZw'$ ).

**Atom:** from  $Z$ 's definition, states  $w$  and  $w'$  agree in all atoms.

**Zig:** Suppose  $wZw'$  and  $w \sim v$ . For a contradiction, suppose no state  $v' \in W'$  satisfies  $w' \sim' v'$  and  $vZv'$ . Define  $S' := \{u' \in W' \mid w' \sim' u'\}$  and note how the set is nonempty. This is because the state  $v$  satisfies  $w \sim v$ , so  $\mathcal{M}, w \models \neg K\perp$ . From this,  $wZw'$  and the definition of  $Z$ , we have  $\mathcal{M}', w' \models \neg K\perp$ ; by  $K$ 's semantic interpretation, it follows that  $w'$  has at least one  $\sim'$ -successor. Now, since  $W'$  is finite,  $S'$  must be finite, so we can write it as  $\{u'_1, \dots, u'_n\}$ . We assumed no state  $u'_i \in S'$  satisfies  $vZu'_i$ , so for each  $u'_i$  there is a formula  $\psi_i$  such that  $\mathcal{M}, v \not\models \psi_i$  and  $\mathcal{M}', u'_i \models \psi_i$ . Thus,  $\mathcal{M}, w \not\models K(\psi_1 \vee \dots \vee \psi_n)$  but  $\mathcal{M}', w' \models K(\psi_1 \vee \dots \vee \psi_n)$ , which contradicts the assumption that  $wZw'$ . Therefore,  $Z$  satisfies **Zig**.

**Zag:** symmetric to the **Zig** clause.

**Khs-Zig:** Suppose  $wZw'$  and  $w \rightsquigarrow U$  for some set  $U \subseteq W$ . Aiming for a contradiction, assume that there is no strategy  $\sigma'$  and no set  $U' \subseteq W'$  such that both  $w' \xrightarrow{\sigma'} U'$  and  $U' \subseteq Z[U]$ . Call this assumption  $(\otimes)$ .

Define the set  $S' := \{U' \subseteq W' \mid w' \xrightarrow{\sigma'} U' \text{ for some strategy } \sigma'\}$ . Since  $w' \rightsquigarrow [w']$ , the set  $S'$  is not empty. Moreover,  $S'$  is finite (because  $W'$  is finite), so we can write it as  $\{U'_1, \dots, U'_k\}$ . By the assumption  $(\otimes)$  above, no set  $U'_i \in S'$  satisfies  $U'_i \subseteq Z[U]$ . In other words, for each  $U'_i \in S'$  there is a state  $v'_i \in U'_i$  such that there is no state  $v \in U$  with  $vZv'_i$ . Hence, by definition of  $Z$ , for each state  $v \in U$  there is a formula  $\varphi_i^v$  such that  $\mathcal{M}, v \models \varphi_i^v$  and  $\mathcal{M}', v'_i \not\models \varphi_i^v$ . Now define

$\varphi_i := \bigvee_{v \in U} \varphi_i^v$  (note that  $U$  is finite). It is clear that  $\varphi_i$  holds on all  $v \in U$  but not on  $v'_i$ . Define the formula  $\psi := \bigwedge_{1 \leq i \leq k} \varphi_i$ . Note that  $\psi$  still holds on all  $v \in U$  but not on those selected  $v'_i \in U'_i$ . It follows that  $\mathcal{M}, w \models \text{Khs}(\psi)$  and  $\mathcal{M}', w' \not\models \text{Khs}(\psi)$ . This contradicts the fact that  $wZw'$ . Thus,  $Z$  satisfies **Khs-Zig**.  
**Khs-Zag**: similar to the previous case. □

**§6. Single-step knowing how.** In [46], the authors introduced a multi-agent *knowing-how* logic to discuss distributed knowledge together with a one-step coalition *knowing-how* operator. By restricting ourselves to the single-agent case (for uniformity reasons, and for making the setting closer to the ones already discussed), the language and semantics can be simplified as follows:

**DEFINITION 6.56 (Language).** *The language  $\mathcal{L}_H$  has three modalities: S for the existence of the strategy, K for knowing that, and H for knowing how. More precisely, formulas  $\varphi$  of the language  $\mathcal{L}_H$  are given by the grammar:*

$$\varphi := \perp \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K\varphi \mid S\varphi \mid H\varphi,$$

with  $p \in \Pi$ .

For the reading of the formulas,  $S\varphi$  intuitively says that there is some strategy which, when executable, can guarantee  $\varphi$  (in [46], a valid formula is the T-axiom for knowing how:  $H\varphi \rightarrow S\varphi$ ).

The semantics is also based on bi-dimensional models with both an epistemic relation (for the single-agent case) and action relations. In the original multi-agent setting of [46], each agent chooses her (global) action among her available ones, and a strategy profile for a group of agents is simply a collective action for those agents. If there is only one agent, then she clearly can decide the global action. Moreover, every action profile is executable on every state. However, to bring it closer to our setting, we do not assume this but add the executability condition for H as given below:

**DEFINITION 6.57 (Semantics).** *Let  $\mathcal{M} = \langle W, \Sigma, \sim, R, V \rangle$  be a bi-dimensional relational model, and let  $w \in W$ . The satisfaction relation  $\models$  between the pointed model  $\mathcal{M}, w$  and a formula  $\varphi$  in  $\mathcal{L}_H$  is inductively defined as before for atoms, negation, conjunction, and the know-that modality; for the new modalities,*

$$\begin{aligned} \mathcal{M}, w \models S\varphi & \text{ iff there is } a \in \Sigma \text{ such that } R_a(w) \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}} \\ \mathcal{M}, w \models H\varphi & \text{ iff there is } a \in \Sigma \text{ such that for all } v \sim w \\ & \quad (1) \text{ } a \text{ is executable on } v, \text{ and} \\ & \quad (2) \text{ } R_a(v) \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}. \end{aligned}$$

Note that H only depends on the single-step strategy, i.e., some basic action  $a$ . Moreover, it is crucial to note that, while here  $H\varphi$  does not entail  $HK\varphi$ , in the setting of [21] the  $\mathcal{L}_{Khs}$  formula  $\text{Khs}(\varphi) \rightarrow \text{Khs}(K\varphi)$  is valid (see [45] for more on this).

**DEFINITION 6.58.** *Let  $\mathcal{M} = \langle W, \Sigma, \sim, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', \sim', R', V' \rangle$  be two bi-dimensional relational models for  $\Pi$ . A nonempty relation  $Z \subseteq W \times W'$  is called a  $\mathcal{L}_{Khs}$ -bisimulation between  $\mathcal{M}$  and  $\mathcal{M}'$  if and only if  $wZw'$  implies **Atom**, **Zig** and **Zag** as in Definition 5.53, and*

- **S-Zig**: if  $w R_a v$  for some action  $a \in \Sigma$ , then there is a state  $v'$  and an action  $a' \in \Sigma'$  such that  $w' R_{a'} v'$  and  $vZv'$ ;



- **S-Zag**: if  $w' R_{a'} v'$  for some action  $a' \in \Sigma'$ , then there is a state  $v$  and an action  $a \in \Sigma$  such that  $w R_a v$  and  $v Z v'$ ;
- **H-Zig**: if  $[w] \xrightarrow{a} U$  for some action  $a \in \Sigma$  and some set  $U \subseteq W$ , then  $[w'] \xrightarrow{a'} U'$  for some action  $a' \in \Sigma'$  and some set  $U' \subseteq W'$  satisfying  $U' \subseteq Z[U]$ ; and
- **H-Zag**: if  $[w'] \xrightarrow{a'} U'$  for some action  $a' \in \Sigma'$  and some set  $U' \subseteq W'$ , then  $[w] \xrightarrow{a} U$  for some action  $a \in \Sigma$  and some set  $U \subseteq W$  satisfying  $U \subseteq Z^{-1}[U']$ ;

where  $[w] \xrightarrow{a} U$  is as in Definition 2.14. We define  $\Leftrightarrow_{\mathcal{L}_H}$  as usual.

With the definition of  $\mathcal{L}_H$ -bisimulations at hand, we can show an invariance result, similarly as it is done in Theorem 5.54.

**THEOREM 6.59** ( $\mathcal{L}_H$  Invariance). *Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed bi-dimensional models with  $\mathcal{M} = \langle W, \Sigma, \sim, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', \sim', R', V' \rangle$ . Then,*

$$\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_H} \mathcal{M}', w' \text{ implies } \mathcal{M}, w \equiv_{\mathcal{L}_H} \mathcal{M}', w'.$$

The other way around also holds when the models have a finite domain.

Based on Definition 6.58, we can define its finite approximations (cf. also [12]). This will be useful in the next section for comparing the expressive power of  $\mathcal{L}_{KHS}$  and  $\mathcal{L}_H$ .

**DEFINITION 6.60.** *Let  $\mathcal{M} = \langle W, \Sigma, \sim, R, V \rangle$  and  $\mathcal{M}' = \langle W', \Sigma', \sim', R', V' \rangle$  be two bi-dimensional relational models for  $\Pi$ . Given  $n \in \mathbb{N}$  and states  $w \in W$  and  $s' \in W'$ , define the relation of  $n$ - $\mathcal{L}_H$  bisimilarity ( $\Leftrightarrow_{\mathcal{L}_H}^n$ ) in the following (inductive) way.*

- $w \Leftrightarrow_{\mathcal{L}_H}^0 w'$  if and only if  $V(w) = V'(w')$ .
- $w \Leftrightarrow_{\mathcal{L}_H}^{k+1} w'$  if and only if all of the following hold.
  - $(k + 1)$ -**Zig**: if  $w \sim v$  then there is a state  $v'$  such that  $w' \sim v'$  and  $v \Leftrightarrow_{\mathcal{L}_H}^k v'$ .
  - $(k + 1)$ -**Zag**: symmetric to the previous clause.
  - $(k + 1)$ -**S-Zig**: if  $w R_a v$  for some action  $a \in \Sigma$  then there is a state  $v'$  and an action  $a' \in \Sigma'$  such that  $w' R_{a'} v'$  and  $v \Leftrightarrow_{\mathcal{L}_H}^k v'$ .
  - $(k + 1)$ -**S-Zag**: symmetric to the previous clause.
  - $(k + 1)$ -**H-Zig**: if  $[w] \xrightarrow{a} U$  for some action  $a \in \Sigma$  and some set  $U \subseteq W$ , then  $[w'] \xrightarrow{a'} U'$  for some action  $a' \in \Sigma'$  and some set  $U' \subseteq W'$  such that for every state  $v' \in U'$  there is a state  $v \in U$  satisfying  $v \Leftrightarrow_{\mathcal{L}_H}^k v'$ .
  - $(k + 1)$ -**H-Zag**: symmetric to the previous clause.

Let  $\equiv_{\mathcal{L}_H}^n$  be the relation of  $\mathcal{L}_H$ -equivalence w.r.t. formulas of modal depth up to  $n$  (counting all the three modalities). Then an analogue of Theorem 6.59 can be obtained:

**THEOREM 6.61** ( $\mathcal{L}_H$  Invariance up-to  $n$ ). *Let  $\mathcal{M}, w$  and  $\mathcal{M}', w'$  be two pointed bi-dimensional models. Then,*

$$\mathcal{M}, w \Leftrightarrow_{\mathcal{L}_H}^n \mathcal{M}', w' \text{ implies } \mathcal{M}, w \equiv_{\mathcal{L}_H}^n \mathcal{M}', w'.$$

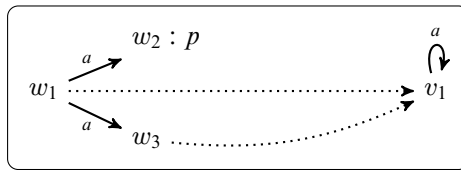
All three modalities S, K, H in  $\mathcal{L}_H$  are one-step in nature. Then, the modal depth of each formula in  $\mathcal{L}_H$  determines its visibility ‘radius’ (the evaluation point being the center), and thus the part of the model it can see.

**6.1. Expressive power of  $\mathcal{L}_{Khs}$  and  $\mathcal{L}_H$ .** As mentioned before, it is not possible to compare the expressive power of the global *knowing-how* logics of Sections 2–4, with that of the bi-dimensional setting of Section 5 and 6. In this section, we will compare the expressivity of the two latter languages.

Intuitively, the *knowing how* operator in  $\mathcal{L}_{Khs}$  is more general than the H operator, which can be viewed as a restriction on single-step strategies. However,  $\mathcal{L}_H$  has a (one-step) strategy operator S, which can say something  $\mathcal{L}_{Khs}$  cannot express.

**THEOREM 6.62.**  *$\mathcal{L}_{Khs}$  and  $\mathcal{L}_H$  are incomparable.*

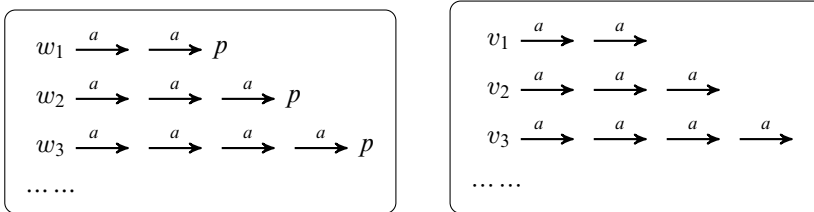
*Proof.* To see that  $\mathcal{L}_H \not\leq \mathcal{L}_{Khs}$ , note that  $S\neg p$  is true at state  $v_1$  but not at state  $w_1$  below (the epistemic relation is the collection of reflexive arrows only, which are omitted from the picture). However,  $w_1$  and  $v_1$  are  $\mathcal{L}_{Khs}$ -bisimilar (witnessed by the dotted lines). As before, dotted lines represent the bisimulation relation.



On the other hand, it is not that straightforward to show  $\mathcal{L}_{Khs} \not\leq \mathcal{L}_H$ , although intuitively the Khs operator in  $\mathcal{L}_{Khs}$  looks more general. The tricky part comes with the coding of Khs by iterating H *given a model*. For example,  $Khs\neg p$  and  $HHK\neg p$  both hold at  $w_1$  in Example 5.50, where the latter is equivalent to the earlier *in that pointed model*.

We need to find two *classes* of (finite) models such that a  $\mathcal{L}_{Khs}$ -formula can distinguish the two classes but given any bound on the modal depth of the  $\mathcal{L}_H$  formulas, there are some pairs of models in those two classes that cannot be distinguished.

We consider the following two classes of models (the epistemic relation is the collection of reflexive arrows only):

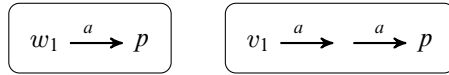


Clearly  $Khs p$  can distinguish each  $w_k$  and  $v_k$  by using a proper strategy which stops at the  $p$ -state. However, for each  $\mathcal{L}_H$ -formula  $\varphi$  of modal depth  $k$  there are  $w_{k+1}$  and  $v_{k+1}$ , which are not distinguishable by  $\varphi$  due to Theorem 6.61. □

Interested readers may wonder what happens if we drop the operator S from  $\mathcal{L}_H$  (we denote this fragment  $\mathcal{L}_H - S$ ). The following result states a comparison between such fragment, and the language  $\mathcal{L}_{Khs}$ .

**THEOREM 6.63.** *The languages  $\mathcal{L}_{Khs}$  and  $\mathcal{L}_H - S$  are still incomparable.*

*Proof.* By using the second example in the proof of Theorem 6.62, we can show that  $\mathcal{L}_{\text{Khs}} \not\leq \mathcal{L}_{\text{H-S}}$ . To show  $\mathcal{L}_{\text{H-S}} \not\leq \mathcal{L}_{\text{Khs}}$ , consider the following two models (the epistemic relations are merely reflexive arrows which are omitted):



It is clear that  $\text{Hp}$  holds at  $w_1$  but not at  $v_1$  since there is no single-step plan. However, it is not hard to show that  $w_1$  and  $v_1$  are  $\mathcal{L}_{\text{Khs}}$ -bisimilar. □

**§7. Related work.** The bisimulation notions proposed in this article have been inspired by monotonic neighbourhood modal logic [5, 7, 27–29, 49, 51, 54]. This is because they coincide with the discussed *knowing how* operators in their use of the  $\exists\forall$  schema suggested by the *de re* reading of ‘*knowing how to achieve  $\varphi$* ’. Bisimulations for other logical systems with this quantification pattern can be found in [2]. A more general notion of bisimulation, for a modal predicate language with a general  $\exists x\Box$  modal operator, can be found in [60].

Several other approaches are also closely related to the work presented in this article. One example is the bisimulation for conditionals studied by [9]. The proposal is based on a framework with formulas of the form  $\psi \rightsquigarrow \varphi$  (‘ $\varphi$  is the case conditional on  $\psi$ ’), semantically interpreted in terms of a *selection function*  $f : W \times \wp(W) \rightarrow \wp(W)$  by stating that  $\psi \rightsquigarrow \varphi$  holds at  $w$  if and only if  $f(w, \llbracket \psi \rrbracket^{\mathcal{M}}) \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}$ . The settings have some similarities, as both  $\psi \rightsquigarrow \varphi$  and  $\text{Kh}(\psi, \varphi)$  guarantee that  $\varphi$  will hold when  $\psi$  is satisfied. Still there are also important differences. First,  $\psi \rightsquigarrow \varphi$  works locally (given a fixed condition  $\psi$ , the relevant set of states might vary depending on the evaluation point), different from the global character of  $\text{Kh}(\psi, \varphi)$ . Moreover, while in the conditional setting, the set  $f(w, \llbracket \psi \rrbracket^{\mathcal{M}})$  is the only relevant one, in the discussed *knowing how* frameworks the relevant sets are all those that can be reached from  $\llbracket \psi \rrbracket^{\mathcal{M}}$  via *some* sequence of actions (that is, the sets in  $\{V \subseteq W \mid \llbracket \psi \rrbracket^{\mathcal{M}} \xrightarrow{\sigma} V \text{ for some sequence } \sigma \in \Sigma^*\}$ ). Thus, the just described conditional setting can be seen as the ‘local’ and ‘single-step’ case of the *knowing how* framework.

Finally, one can find connections with bisimulations for logics with strategies such as ATL [4], and in particular with ATL logics with knowledge operators [55]. However, as mentioned in the introduction, a simple combination of knowledge and ability does not capture *knowing how* properly [31]. Various alternative semantics have been proposed in the literature (e.g., [15]). For example, in [10] a bisimulation notion for ATL was studied under *subjective semantics*. The crucial idea behind their Zig-Zag conditions is similar to ours: for each strategy  $\sigma$  in one model there is a corresponding strategy  $\sigma'$  in the other model such that any point in the outcome set for  $\sigma'$  is linked by the bisimulation to a point in the outcome set for  $\sigma$ . In a more recent work, in [11] bisimulation notions for different fragments of strategy logic were introduced, using them to compare the relative expressive power of such fragments. In particular, the authors consider a multi-agent setting that includes coalition strategies, more related to the setting in [46].

**§8. Final remarks.** In this article, we introduced suitable notions of bisimulation for five different *knowing how* logics. The semantics of the first *knowing how* operator

(first introduced in [59, 62]) requires plans that always succeed in order to know how to achieve  $\varphi$  given a precondition  $\psi$ . The second alternative (from [38]) incorporates conditions in each intermediate state that is ‘visited’ while executing the successful plan. A weaker *knowing how* operator was introduced in [37], in which it is considered that the goal  $\varphi$  is achieved when it holds in all states that can be reached via the given plan, even those in which the plan cannot be finished. Finally, we consider the knowing how operator from [21] and a variant from [46], that combine local knowing how and knowing that modalities, allowing the execution of conditional plans.

As we pointed out, these operators were already investigated but the model theoretical aspects have not been studied before. For the five notions of bisimulation introduced, we proved that bisimilarity implies modal equivalence in the corresponding logic, and when the class of finite models is considered, modal equivalence implies bisimilarity. We discussed how the definitions of bisimulation we introduced are related with similar notions, such as those for conditional logic [9], ATL [4, 55] and strategy logic [11]. As an application, bisimulations are used as a tool to investigate the relative expressive power of the corresponding logics.

We believe the results presented here are the first step towards a better understanding of the model theory and expressivity of this family of logics. As future work, the algorithms to compute some notion of bisimulation contraction would help the applications of both knowing how logic and epistemic planning (cf. [16]). It would also be interesting to explore other model theoretical properties for this family of logics, which can capture interactions between *knowing how* and *knowing that* operators. Also, it would be interesting to explore the computational complexity of the inference tasks for different knowing how frameworks, and analyze the trade-off between expressivity and good computational behavior. Of course, another extension is to move to the multi-agent setting, in order to describe the *knowing how* abilities of different agents. When combined with a *knowing that* operator as in the just mentioned proposals, this would give us a setting to talk about the knowledge agents have about one another’s abilities.

**Acknowledgments.** We would like to thank the anonymous reviewers for their helpful comments and suggestions. R. Fervari is supported by projects ANPCyT-PICT 2017-1130, Stic-AmSud 20-STIC-03 ‘DyLo-MPC’, Secyt-UNC, GRFT Mincyt-Cba, and by the Laboratoire International Associé SINFIN. Y. Wang acknowledges the support from NSSF grant 19BZX135.

## BIBLIOGRAPHY

- [1] Ågotnes, T. & Alechina, N. (2019). Coalition logic with individual, distributed and common knowledge. *Journal of Logic and Computation*, **29**(7), 1041–1069.
- [2] Ågotnes, T., Goranko, V., & Jamroga, W. (2007). Alternating-time temporal logics with irrevocable strategies. In *Proceedings of Conference on Theoretical Aspects of Rationality and Knowledge*. New York: ACM, pp. 15–24.
- [3] Ågotnes, T., Goranko, V., Jamroga, W., & Wooldridge, M. (2015). Knowledge and ability (see van Ditmarsch et al. (2015), Ch. 11). In *Handbook of Epistemic Logic*. London: College Publications, pp. 543–589.
- [4] Alur, R., Henzinger, T., & Kupferman, O. (2002). Alternating-time temporal logic. *Journal of the ACM*, **49**(5), 672–713.

[5] Areces, C. & Figueira, D. (2009). Which semantics for neighbourhood semantics? In *Proceedings of IJCAI 2009*. San Francisco, CA: AAAI Press. pp. 671–676.

[6] Artemov, S. & Fitting, M. (2019). *Justification Logic: Reasoning with Reasons*. Cambridge Tracts in Mathematics, Vol. 216. Cambridge: Cambridge University Press.

[7] Bakhtiari, Z., van Ditmarsch, H., & Hansen, H. H. (2017). Neighbourhood contingency bisimulation. In *Proceedings of ICLA*. Berlin: Springer, pp. 48–63.

[8] Baltag, A. (2016). To know is to know the value of a variable. In *Proceedings of Advances in Modal Logic*, Vol. 11. London: College Publications, pp. 135–155.

[9] Baltag, A. & Ciná, G. (2018). Bisimulation for conditional modalities. *Studia Logica*, **106**(1), 1–33.

[10] Belardinelli, F., Condurache, R., Dima, C., Jamroga, W., & Jones, A. V. (2017). Bisimulations for verifying strategic abilities with an application to threeballot. In *Proceedings of AAMAS*. New York: ACM, pp. 1286–1295.

[11] Belardinelli, F., Dima, C., & Murano, A. (2018). Bisimulations for logics of strategies: a study in expressiveness and verification. In Thielscher, M., Toni, F., and Wolter, F., editors. *Proceedings of KR*. San Francisco, CA: AAAI Press. pp. 425–434.

[12] Blackburn, P., de Rijke, M., & Venema, Y. (2002). *Modal Logic*. Cambridge: Cambridge University Press.

[13] Blackburn, P. & van Benthem, J. (2006). Modal logic: a semantic perspective. In *Handbook of Modal Logic*. North Holland: Elsevier. pp. 1–84.

[14] Bolander, T. & Andersen, M. B. (2011). Epistemic planning for single and multi-agent systems. *Journal of Applied Non-Classical Logics*, **21**(1), 9–34.

[15] Bulling, N. & Jamroga, W. (2014). Comparing variants of strategic ability: how uncertainty and memory influence general properties of games. *Autonomous Agents and Multi-Agent Systems*, **28**(3), 474–518.

[16] Dovier, A., Piazza, C. & Policriti, A. (2004). An efficient algorithm for computing bisimulation equivalence. *Theoretical Computer Science*, **311**(1–3), 221–256.

[17] Duijf, H. (2018). *Let's Do It!: Collective Responsibility, Joint Action, and Participation*. PhD Thesis, Utrecht University, Utrecht, The Netherlands.

[18] Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning About Knowledge*. Cambridge, MA: The MIT Press.

[19] Fan, J., Wang, Y., & van Ditmarsch, H. (2015). Contingency and knowing whether. *The Review of Symbolic Logic*, **8**, 75–107.

[20] Fantl, J. (2017). Knowledge how. In Zalta, E. N., editor. *The Stanford Encyclopedia of Philosophy* (Fall 2017 edition). Stanford, CA: Metaphysics Research Lab.

[21] Fervari, R., Herzig, A., Li, Y., & Wang, Y. (2017). Strategically knowing how. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. IJCAI, USA: IJCAI, pp. 1031–1038.

[22] Fervari, R., Velázquez-Quesada, F. R., & Wang, Y. (2017). Bisimulations for knowing how logics. Presented at 5th International Workshop on Strategic Reasoning. Liverpool, UK.

[23] Gochet, P. (2013). An open problem in the logic of knowing how. In Hintikka, J., editor. *Open Problems in Epistemology*. Helsinki: The Philosophical Society of Finland.

- [24] Goranko, V. & Otto, M. (2006). Model theory of modal logic. In *Handbook of Modal Logic*, Vol. 3. New York: Elsevier Science, Inc., pp. 249–329.
- [25] Goranko, V. & Passy, S. (1992). Using the universal modality: gains and questions. *Journal of Logic and Computation*, **2**(1), 5–30.
- [26] Gu, T. & Wang, Y. (2016). “knowing value” logic as a normal modal logic. In *Proceedings of Advances in Modal Logic*, Vol. 11. London: College Publications, pp. 362–381.
- [27] Hansen, H. H. (2003). *Monotonic modal logics*. Master’s Thesis, Universiteit van Amsterdam.
- [28] Hansen, H. H., Kupke, C., & Pacuit, E. (2007). Bisimulation for neighbourhood structures. In Mossakowski T., Montanari U., and Haveraaen M., editors. *Algebra and Coalgebra in Computer Science*. Lecture Notes in Computer Science, Vol. 4624. Berlin: Springer, pp. 279–293.
- [29] ———. (2009). Neighbourhood structures: bisimilarity and basic model theory. *Logical Methods in Computer Science*, **5**(2), 1–38.
- [30] Hart, S., Heifetz, A., & Samet, D. (1996). Knowing whether, knowing that, and the cardinality of state spaces. *Journal of Economic Theory*, **70**(1), 249–256.
- [31] Herzig, A. (2015). Logics of knowledge and action: critical analysis and challenges. *Autonomous Agents and Multi-Agent Systems*, **29**(5), 719–753.
- [32] Herzig, A. & Troquard, N. (2006). Knowing how to play: uniform choices in logics of agency. In *Proceedings of AAMAS*. New York: ACM, pp. 209–216.
- [33] Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, NY: Cornell University Press.
- [34] Hodges, W. (1993). *Model Theory*. Encyclopedia of Mathematics and Its Applications, Vol. 42. Cambridge: Cambridge University Press.
- [35] Horty, J. F. & Pacuit, E. (2017). Action types in STIT semantics. *Review of Symbolic Logic*, **10**(4), 617–637.
- [36] Jamroga, W. & Ågotnes, T. (2007). Constructive knowledge: what agents can achieve under imperfect information. *Journal of Applied Non-Classical Logics*, **17**(4), 423–475.
- [37] Li, Y. (2017). Stopping means achieving: a weaker logic of knowing how. *Studies in Logic*, **9**(4), 34–54.
- [38] Li, Y. & Wang, Y. (2017). Achieving while maintaining: a logic of knowing how with intermediate constraints. In *Proceedings of ICLA*. Heidelberg: Springer, pp. 154–167.
- [39] ———. (2019). Multi-agent knowing how via multi-step plans: a dynamic epistemic planning based approach. In *Proceedings of LORI 2019*. Berlin: Springer Nature, pp. 126–139.
- [40] Li, Y., Yu, Q., & Wang, Y. (2017). More for free: a dynamic epistemic framework for conformant planning over transition systems. *Journal of Logic and Computation*, **27**(8), 2383–2410.
- [41] Meyer, J.-J. C. & van Der Hoek, W. (1995). *Epistemic Logic for AI and Computer Science*. New York: Cambridge University Press.
- [42] Naumov, P. & Tao, J. (2017a). Coalition power in epistemic transition systems. In *Proceedings of AAMAS*. New York: ACM, pp. 723–731.
- [43] ———. (2017b). Together we know how to achieve: an epistemic logic of know-how (extended abstract). In *Proceedings of Conference on Theoretical Aspects of Rationality and Knowledge*. Australia: Open Publishing Association, pp. 441–453.



- [44] ———. (2018a). Second-order know-how strategies. In *Proceedings of AAMAS*, Richland, SC: IFAAMAS, pp. 390–398.
- [45] ———. (2018b). Strategic coalitions with perfect recall. In *Proceedings of AAAI*. San Francisco, CA: AAAI Press, pp. 4702–4709.
- [46] ———. (2018c). Together we know how to achieve: an epistemic logic of know-how. *Artificial Intelligence*, **262**, 279–300.
- [47] ———. (2019). Knowing-how under uncertainty. *Artificial Intelligence*, **276**, 41–56.
- [48] Nissim, R., Hoffmann, J., & Helmert, M. (2011). Computing perfect heuristics in polynomial time: on bisimulation and merge-and-shrink abstraction in optimal planning. In *Proceedings of IJCAI 2011*. Menlo Park, CA: AAAI, pp. 1983–1990.
- [49] Pacuit, E. (2017). *Neighborhood Semantics for Modal Logic*. Berlin: Springer.
- [50] Parikh, R. & Ramanujam, R. (2003). A knowledge based semantics of messages. *Journal of Logic, Language and Information*, **12**(4), 453–467.
- [51] Pauly, M. (1999). Bisimulation for general non-normal modal logic. Unpublished manuscript.
- [52] Sangiorgi, D. (2011). *Introduction to Bisimulation and Coinduction*. Cambridge: Cambridge University Press.
- [53] Stanley, J. & Williamson, T. (2001). Knowing how. *The Journal of Philosophy*, **98**(8), 411–444.
- [54] van Benthem, J., Bezhanishvili, N., Enqvist, S., & Yu, J. (2017). Instantial neighbourhood logic. *The Review of Symbolic Logic*, **10**(1), 116–144.
- [55] van der Hoek, W. & Wooldridge, M. (2003). Cooperation, knowledge, and time: alternating-time temporal epistemic logic and its applications. *Studia Logica*, **75**(1), 125–157.
- [56] van Ditmarsch, H., Halpern, J. Y., van der Hoek, W., & Kooi, B., editors. (2015). *Handbook of Epistemic Logic*. London: College Publications.
- [57] van Eijck, J., Gattinger, M., & Wang, Y. (2017). Knowing values and public inspection. In *Proceedings of ICLA*. Berlin, pp. 77–90.
- [58] Wang, X. (2019). A logic of knowing how with skippable plans. In *Proceedings of LORI*. Berlin: Springer Nature, pp. 413–424.
- [59] Wang, Y. (2015). A logic of knowing how. *Proceedings of LORI*. Lecture Notes in Computer Science, Vol. 9394. Berlin: Springer, pp. 392–405.
- [60] ———. (2017). A new modal framework for epistemic logic. In *Proceedings of Conference on Theoretical Aspects of Rationality and Knowledge*. Australia: Open Publishing Association, pp. 515–534.
- [61] ———. (2018a). Beyond knowing that: a new generation of epistemic logics. In van Ditmarsch, H. & Sandu, G., editors. *Jaakko Hintikka on Knowledge and Game Theoretical Semantics*. Berlin: Springer, pp. 499–533.
- [62] ———. (2018b). A logic of goal-directed knowing how. *Synthese*, **195**(10), 4419–4439.
- [63] Wang, Y. & Li, Y. (2012). Not all those who wander are lost: dynamic epistemic reasoning in navigation. In *Advances in Modal Logic*, Vol. 10. pp. 559–580.
- [64] Xu, C., Wang, Y., & Studer, T. (2021). A logic of knowing why. *Synthese*, **198**, 1259–1285.



FACULTAD DE MATEMÁTICA, ASTRONOMÍA, FÍSICA Y COMPUTACIÓN  
UNIVERSIDAD NACIONAL DE CÓRDOBA  
MEDINA ALLENDE S/N, CÓRDOBA, ARGENTINA

*E-mail:* [rfervari@unc.edu.ar](mailto:rfervari@unc.edu.ar)

*URL:* <http://cs.famaf.unc.edu.ar/~rfervari>

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION  
UNIVERSITEIT VAN AMSTERDAM  
P.O. BOX 94242, 1090, GE AMSTERDAM, THE NETHERLANDS

*E-mail:* [F.R.VelazquezQuesada@uva.nl](mailto:F.R.VelazquezQuesada@uva.nl)

*URL:* <http://staff.fnwi.uva.nl/f.r.velazquezquesada>

DEPARTMENT OF PHILOSOPHY  
PEKING UNIVERSITY  
100871 BEIJING, CHINA

*E-mail:* [y.wang@pku.edu.cn](mailto:y.wang@pku.edu.cn)

*URL:* <http://wangyanjing.com/>