Bisimulations for *Knowing How* Logics*

Raul Fervari
Universidad Nacional de Córdoba and CONICET, Argentina
rfervari@conicet.gov.ar

Fernando R. Velázquez-Quesada
Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands
F.R.VelazquezQuesada@uva.nl

Yanjing Wang
Department of Philosophy, Peking University, China
y.wang@pku.edu.cn

This paper introduces suitable notions of bisimulation for a family of logics of *knowing how*, which capture the epistemic reasoning about goal-directed know-how by a single agent. First, we present a bisimulation notion for the basic *knowing how* logic. We prove that with this notion, bisimilarity implies formula equivalence, and that over finite models the converse also holds. Then, we explore bisimulations for *knowing how* logic with intermediate states, and for a weaker *knowing how* logic.

1 Introduction

Standard epistemic logic [15, 8] studies the reasoning patterns of *knowing that*. It has been successfully applied to various fields, such as theoretical computer science, game theory, and AI (cf. e.g., [8, 19, 7]). However, *knowing that* is not the only interesting kind of knowledge. For example, in strategic reasoning and automated planning, it is especially important to ask whether an agent knows *how* to achieve certain goals (cf. e.g., the surveys [2, 10, 24]). Interestingly, *knowing how* cannot be captured by a simple combination of *knowing that* and ability, as shown in [16, 14]. This is because a philosophically grounded formulation of *knowing how to achieve* φ requires a *de re* reading: there exists a method x such that the agent knows that executing x guarantees φ [22]. Also important is to take a *global* perspective: the given strategy should allow the agent to achieve the goal *in every scenario satisfying the initial conditions*, in order to assure that the success is a matter of ability and not a matter of luck. Based on this reading, Wang proposed and studied a logic featuring a simple goal-directed *knowing how* operator Kh [23, 25], of which variants have been studied in [18, 17, 9]. These logics are not normal, e.g., knowing how to get drunk and knowing how to prove a theorem do not imply knowing how to get the theorem proved and drunk at the same time. Moreover, given the semantic interpretation of their *knowing how* operators, these languages can define the universal modality and have characteristic axioms capturing the composition of plans or strategies. All this makes the family of *knowing how* logics interesting in the eyes of logicians.

Previous works have mainly focussed on axiomatizations, and thus have not thoroughly studied the model theoretical aspects of these logics. In this paper, we make the first step in this direction by studying bisimulation notions for three *knowing how* logics: the basic one in [25], the one with intermediate constraints in [18], and the weakly *knowing how* logic in [17]. We obtain invariance results and prove Hennessy-Milner-style theorems over finite models. These bisimulation notions, which can be used to minimize models, also sharpen our understanding of the expressivity of those logics.

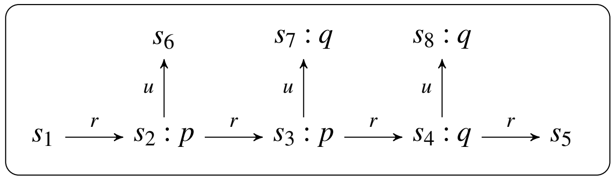
We use an example from [27] to illustrate the intuitive differences among the three *knowing how* operators corresponding to each *knowing how* logic mentioned above. In this context, relational models as in the example can be viewed as a reflection of the abilities of the agent, thus they are called *ability maps*, which induce the knowledge-how of the agent.[1]

*Extended abstract

[1] Note that the following *knowing how* operators are global (cf. [25] for discussions).

Submitted to:
SR 2017 - Liverpool, UK

Example 1. *Consider the following ability map:*



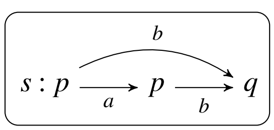
Let us see the differences considering each definition of knowing how.

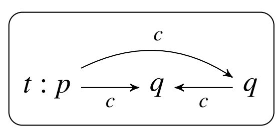
1. *Note how the agent knows how to reach q given s(he) is at a p-state (in symbols, the $\mathsf{Kh}(p,q)$ proposed in [23, 25] holds). This is because there exists a uniform linear plan (ru) such that from* any *p-state (s_2 and s_3) the agent can always* successfully *execute the plan and ensure to reach some q-state (s_7 and s_8, respectively).*
2. *On the other hand, the agent does not know how to reach q given p by* only passing through *p-states (in symbols, the $\mathsf{Khm}(p,p,q)$ discussed in [18, 25] fails). The only uniform plan that guarantees q given p, namely ru, has to travel through a $\neg p$-state (s_4) if the agent starts at s_3.*
3. *Finally, the agent* weakly *knows how to reach $\neg p \wedge \neg q$ given p (in symbols, the $\mathsf{Kh}^W(p,\neg p \wedge \neg q)$ studied in [17] holds), since there exists a uniform plan rrr that always reaches a $(\neg p \wedge \neg q)$-state* whenever it can be executed. *In particular, note how rrr cannot be successfully executed at s_3: after executing rr, the agent will reach s_5, from where r can no longer take place. Thus, rrr is not a suitable plan for the stronger* knowing how *of the first item in this list. (In fact, this stronger $\mathsf{Kh}(p,\neg p \wedge \neg q)$ fails, as there is no suitable uniform executable plan that can be successfully executed from any p-state leading in every case to a $(\neg p \wedge \neg q)$-state.)*

The condition in the above *knowing how* operators naturally induces an initial epistemic uncertainty, e.g., $\mathsf{Kh}(p,q)$ can be read as the agent knows how to reach q given (s)he only knows p.

In order to find suitable bisimulation notions, we need to overcome one technical difficulty shared by all these logics: the imbalance between weak languages and rich models. Note that while the logic in [25] has a single *knowing how* modality, it is interpreted over arbitrary labelled transition systems. Thus, a simple adaptation of the standard bisimulation for modal logic does not work.

Example 2. *The following two pointed models w.r.t. s and t respectively satisfy exactly the same* Kh-*formulas, but they are not bisimilar at all w.r.t. the usual notion of bisimulation over transition systems:*





Our crucial inspiration comes from monotonic neighbourhood modal logic [21, 11, 12, 3, 13], which also shares the $\exists\forall$ schema in its semantics.[2] We may look at the transition systems from a very abstract point of view, and only consider the "forcing relation" taking the agent from a given set of states U to those states the agent can reach by the execution of some plan/strategy from U.[3] A more general notion of bisimulation for the predicate language with a general $\exists x \Box$ modal operator can be found in [26].

The paper starts by reviewing the definitions of the basic *knowing how* logic (§2), and then proposes and studies an appropriate bisimulation notion (§3). The ideas are extended for dealing with two other *knowing how* logics (§4 and §5) before concluding with future directions (§6).

[2]Other systems also make use of this quantification pattern; for bisimulations on some of them, see, e.g., [1].
[3]See [4, Ch. 11] for a detailed discussion on similar forcing relations in the setting of games.

2 Basic Definitions

Through the text, let PROP be a countable set of propositional symbols, and let Σ be a countable non-empty set of action symbols.

Definition 1 (Syntax)**.** *We define the set of $\mathscr{L}_{\mathsf{Kh}}$-formulas as:*

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

with $p \in \mathsf{PROP}$ (note how no formula in $\mathscr{L}_{\mathsf{Kh}}$ refers to actions in Σ). Formulas of the form $\mathsf{Kh}(\psi, \varphi)$ express that, given ψ, the agent knows how to achieve φ. Other Boolean constants and connectives ($\vee$, $\top$, $\rightarrow$, $\leftrightarrow$) are defined as usual. Additionally, define $\mathsf{A}\varphi := \mathsf{Kh}(\neg\varphi, \bot)$ and $\mathsf{E}\varphi := \neg\mathsf{A}\neg\varphi$. Finally, $\mathscr{L}$ denotes the propositional fragment of $\mathscr{L}_{\mathsf{Kh}}$.

Definition 2 (Relational Models)**.** *A* relational model *$\mathscr{M}$ (also called* transition systems *and* ability maps *in the literature) is a tuple $\langle W, \mathrm{R}, \mathrm{V}\rangle$ where W is a non-empty set of possible states (sometimes denoted by $D(\mathscr{M})$), $\mathrm{R} : \Sigma \rightarrow \wp(W \times W)$ assigns a binary relation on W to each $a \in \Sigma$, and $\mathrm{V} : W \rightarrow \wp(\mathsf{PROP})$ is an atomic valuation. A pair $(\mathscr{M}, w)$ with $\mathscr{M}$ a relational model and $w \in D(\mathscr{M})$ is called a* pointed (relational) model*, with w its* evaluation point *(we usually drop parentheses).*

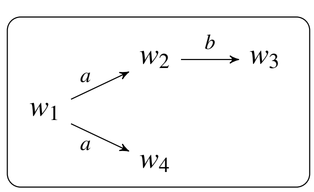
The following definitions will be useful.

Definition 3. *Let $\mathscr{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ be a relational model.*

- *Take $a \in \Sigma$. We write $w \xrightarrow{a} v$ whenever $(w, v) \in \mathrm{R}(a)$.*
- *Let $\sigma = a_1 \cdots a_n \in \Sigma^*$ be a sequence. Given $w, v \in W$, we write $w \xrightarrow{\sigma} v$ when there are $u_1, \ldots, u_{n-1} \in W$ such that*
  $$w \xrightarrow{a_1} u_1 \xrightarrow{a_2} \ldots \xrightarrow{a_{n-1}} u_{n-1} \xrightarrow{a_n} v$$
  Note that σ can be the empty sequence ε, which satisfies $w \xrightarrow{\varepsilon} w$ for any $w \in W$.
- *We say that σ is* executable *at $w \in W$ if and only if there is $v \in W$ such that $w \xrightarrow{\sigma} v$.*
- *Given a sequence $\sigma = a_1 \cdots a_n \in \Sigma^*$ and a $k \leq |\sigma|$, denote by σ_k the initial segment of σ up to a_k (i.e., $\sigma_k := a_1 \cdots a_k$). In particular, $\sigma_0 := \varepsilon$.*
- *We say that $\sigma = a_1 \ldots a_n$ is* strongly executable (s.e.) *at w if and only if: for any $0 \leq k < n$ and any $v \in W$, $w \xrightarrow{\sigma_k} v$ implies that v has at least one a_{k+1}-successor. It is not hard to see that, if σ is strongly executable at w, then it is executable at w.*
- *Given $\sigma \in \Sigma^*$ and $U \subseteq W$, define $\mathrm{R}_\sigma[U] := \{v \in W \mid u \xrightarrow{\sigma} v \text{ for some } u \in U\}$, so $\mathrm{R}_\sigma[U]$ is the set of states reachable via σ from some element of U.*
- *We write $U \xrightarrow{\sigma} V$ whenever σ is s.e. for all $u \in U$ and $V = \mathrm{R}_\sigma[U]$.*
- *We write $U \rightarrow V$ whenever there is a $\sigma \in \Sigma^*$ such that $U \xrightarrow{\sigma} V$.*

It is important to notice the difference between executability and strong executability.

Example 3. *Consider the following relational model.*



Note how the sequence ab is executable at w_1, as there are w_2 and w_3 such that $w_1 \xrightarrow{a} w_2 \xrightarrow{b} w_3$. However, ab is not *strongly executable at w_1: for its subsequence a we have $w_1 \xrightarrow{a} w_4$, but nevertheless w_4 does not have a b-successor.*

Definition 4 (Semantics)**.** *Let $\mathscr{M} = \langle W, R, V \rangle$ be a model, and take $w \in W$. The satisfaction relation $\models$ is defined as follows:*

$$\begin{array}{lll}
\mathscr{M}, w \not\models \bot & & \\
\mathscr{M}, w \models p & \textit{iff} & p \in V(w) \\
\mathscr{M}, w \models \neg\varphi & \textit{iff} & \mathscr{M}, w \not\models \varphi \\
\mathscr{M}, w \models \varphi \wedge \psi & \textit{iff} & \mathscr{M}, w \models \varphi \textit{ and } \mathscr{M}, w \models \psi \\
\mathscr{M}, w \models \mathsf{Kh}(\psi, \varphi) & \textit{iff} & \textit{there exists } \sigma \in \Sigma^* \textit{ such that for all } u \in W, \textit{ if } \mathscr{M}, u \models \psi, \textit{ then} \\
& & \sigma \textit{ is s.e. at } u \textit{ and, for all } v \in W,\ u \xrightarrow{\sigma} v \textit{ implies } \mathscr{M}, v \models \varphi.
\end{array}$$

We say φ is satisfiable *if and only if there exists a pointed model $\mathscr{M}, w$ such that $\mathscr{M}, w \models \varphi$. We say φ is* valid *if and only if $\mathscr{M}, w \models \varphi$ for every pointed model $\mathscr{M}, w$. Additionally, we define $[\![\varphi]\!]^{\mathscr{M}} = \{w \mid \mathscr{M}, w \models \varphi\}$, the extension of φ in the model $\mathscr{M}$.*

Then, at state w of model $\mathscr{M}$ the agent knows how to achieve φ given ψ, $\mathscr{M}, w \models \mathsf{Kh}(\psi, \varphi)$, if and only if there is a sequence σ that is strongly executable from any ψ-state, and any of such executions leads to a φ-state. Note how the evaluation point w does not play any role in the semantic clause, which focusses rather on global model properties. Moreover, $[\![\mathsf{Kh}(\psi, \varphi)]\!]^{\mathscr{M}}$ is either the full domain or the empty set. Then, given that for every $U \subseteq D(\mathscr{M})$ we have $U \xrightarrow{\varepsilon} U$, the semantic interpretation of the abbreviation $\mathsf{A}\varphi$ is such that

$$\mathscr{M}, w \models \mathsf{A}\varphi \quad \text{iff} \quad \text{for all } u \in W, \text{ we have } \mathscr{M}, u \models \varphi$$

Thus, A is the *global* universal modality (and hence E is its *global* existential dual).

For a sound and complete axiomatization of validities in $\mathscr{L}_{\mathsf{Kh}}$ over the class of relational models, the reader is referred to [23]. We finish this section with two concepts: $\mathscr{L}_{\mathsf{Kh}}$-equivalence and definability.

Definition 5 ($\mathscr{L}_{\mathsf{Kh}}$-equivalence)**.** *Let $\mathscr{M}, w$ and $\mathscr{M}', w'$ be two pointed models. We say $\mathscr{M}, w$ and $\mathscr{M}', w'$ are* $\mathscr{L}_{\mathsf{Kh}}$-equivalent *(notation: $\mathscr{M}, w \equiv_{\mathscr{L}_{\mathsf{Kh}}} \mathscr{M}', w'$) if and only if they satisfy the same $\mathscr{L}_{\mathsf{Kh}}$-formulas, i.e., if and only if, for all φ in $\mathscr{L}_{\mathsf{Kh}}$, we have $\mathscr{M}, w \models \varphi$ iff $\mathscr{M}', w' \models \varphi$.*

Definition 6 (Definability)**.** *Let $\mathscr{M}$ be a relational model. A set $U \subseteq D(\mathscr{M})$ is* $\mathscr{L}_{\mathsf{Kh}}$-definable (resp., $\mathscr{L}$-definable, or *propositionally definable*) in $\mathscr{M}$ *if and only if there is $\varphi \in \mathscr{L}_{\mathsf{Kh}}$ (resp., $\varphi \in \mathscr{L}$) such that $U = [\![\varphi]\!]^{\mathscr{M}}$.*

3 Bisimulation

In this section we introduce a notion of bisimulations for the *knowing how* language $\mathscr{L}_{\mathsf{Kh}}$. As it will be proved, this notion provides (in finite models) a semantic characterization of the language's expressivity, and thus it allows us to compare $\mathscr{L}_{\mathsf{Kh}}$ with other proposals.

When looking for a notion of bisimulation for a given language, one needs to be careful when formulating the conditions: they should be strong enough so that they guarantee 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 many cases, the definition will look natural after being provided, but the conditions might not have been obvious at all before.

Let $\mathscr{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ and $\mathscr{M}' = \langle W', \mathrm{R}', \mathrm{V}'\rangle$ be two relational models. It is clear that the standard bisimulation for the basic multi-modal language is not adequate, as the languages have different expressivity (for one, the global modality is not definable within the basic multi-modal language, but it is definable within $\mathscr{L}_{\mathsf{Kh}}$). Still, there are some basic ingredients that should appear. From the fact that the language $\mathscr{L}_{\mathsf{Kh}}$ involves atomic propositions and Boolean operators, it is obvious that an appropriate bisimulation relation $Z \subseteq W \times W'$ should only contain pairs of worlds with matching atomic valuation.

Now, for the *Zig* and *Zag* (also called *forth* and *back*) conditions. They should be designed to match the lone 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 emphasised, the actual evaluation point does not play any role in the semantic interpretation of the knowing how modality. In fact, as the global modality is definable in $\mathscr{L}_{\mathsf{Kh}}$, every world in W should have a 'matching' world in W', and vice-versa, regardless of whether they are accessible from the given pair (w, w').

But this is, of course, not enough: the actual relations between worlds have not been considered, 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 bisimilarity indeed implies $\mathscr{L}_{\mathsf{Kh}}$-equivalent.

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 $\Box$ 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 $\mathscr{L}_{\mathsf{Kh}}$ requires can be found in the literature for bisimulations *over neighbourhood models* (e.g., [21, 11, 12, 3, 13]).

Definition 7. *Let $\mathscr{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ and $\mathscr{M}' = \langle W', \mathrm{R}', \mathrm{V}'\rangle$ be two relational models. A non-empty relation $Z \subseteq W \times W'$ is called an $\mathscr{L}_{\mathsf{Kh}}$-bisimulation between $\mathscr{M}$ and $\mathscr{M}'$ if and only if wZw' implies:*

Atom: $\mathrm{V}(w) = \mathrm{V}'(w')$.

Kh-**Zig:** *for any propositionally definable $U \subseteq W$, if $U \to V$ for some $V \subseteq W$, then there is $V' \subseteq W'$ such that (i) $Z[U] \to V'$ and, (ii) for each $v' \in V'$ there is a $v \in V$ such that vZv'.*

Kh-**Zag:** *for any propositionally definable $U' \subseteq W'$, if $U' \to V'$ for some $V' \subseteq W'$, then there is $V \subseteq W$ such that (i) $Z^{-1}[U'] \to V$ and (ii) for each $v \in V$ there is a $v' \in V'$ such that vZv'.*

A-**Zig:** *for all v in W there is a v' in W' such that vZv'.*

A-**Zag:** *for all v' in W' there is a v in W such that vZv'.*

where $Z[U] = \{w' \mid wZw' \text{ for some } w \in U\}$ and $Z^{-1}[U'] = \{w \mid wZw' \text{ for some } w' \in U'\}$. We write $\mathscr{M}, w \rightleftharpoons_{\mathscr{L}_{\mathsf{Kh}}} \mathscr{M}', w'$ when there is an $\mathscr{L}_{\mathsf{Kh}}$-bisimulation Z between $\mathscr{M}$ and $\mathscr{M}'$ such that wZw'.

This definition of bisimulation has at least two points worthwhile to expand on. The first is about the Kh-Zig and Kh-Zag clauses. Different from the standard bisimulation for the basic modal language (see, e.g., Section 2.2 of [6]), they do not rely on worlds accessible from the pair (w, w') under evaluation; they rather work with arbitrary *propositionally definable* subsets of the domain. That they act globally over sets of worlds is natural, given the semantic interpretation of the lone modal operator of the language, Kh. The requirement of definable sets is also reasonable, as we are only interested in sets of worlds the language can distinguish. Finally, the restriction to *propositionally* definable sets might seem too strong, but this is actually not the case: any $\mathscr{L}_{\mathsf{Kh}}$-definable set is also propositionally definable.

Proposition 1. *Let $\mathscr{M}$ be a relational model. For all $U \subseteq D(\mathscr{M})$, if U is $\mathscr{L}_{\mathsf{Kh}}$-definable, then it is $\mathscr{L}$-definable.*

Proof sketch. It follows from the fact that the lone non-Boolean operator in $\mathscr{L}_{\mathsf{Kh}}$, the knowing-how operator $\mathsf{Kh}(\psi,\varphi)$, is such that $[\![\mathsf{Kh}(\psi,\varphi)]\!]^{\mathscr{M}}$ is either $D(\mathscr{M})$ or else $\emptyset$, and thus it is semantically equivalent to either $\top$ or else $\bot$. More precisely, one can define a translation $\mathrm{tr}_{\mathscr{M}} : \mathscr{L}_{\mathsf{Kh}} \to \mathscr{L}$ such that, for every $\varphi \in \mathscr{L}_{\mathsf{Kh}}$, $[\![\varphi]\!]^{\mathscr{M}} = [\![\mathrm{tr}_{\mathscr{M}}(\varphi)]\!]^{\mathscr{M}}$. The definition is the natural one for atoms and Boolean constants and connectives; for formulas of the form $\mathsf{Kh}(\psi,\varphi)$, it simply chooses between $\top$ and $\bot$, according to whether $[\![\mathsf{Kh}(\psi,\varphi)]\!]^{\mathscr{M}}$ is $D(\mathscr{M})$ or $\emptyset$:

$$\mathrm{tr}_{\mathscr{M}}(\bot) := \bot \qquad \mathrm{tr}_{\mathscr{M}}(p) := p$$

$$\mathrm{tr}_{\mathscr{M}}(\neg\varphi) := \neg\,\mathrm{tr}_{\mathscr{M}}(\varphi) \qquad \mathrm{tr}_{\mathscr{M}}(\varphi \wedge \psi) := \mathrm{tr}_{\mathscr{M}}(\varphi) \wedge \mathrm{tr}_{\mathscr{M}}(\psi)$$

$$\mathrm{tr}_{\mathscr{M}}(\mathsf{Kh}(\psi,\varphi)) := \begin{cases} \top & \text{if } [\![\mathsf{Kh}(\psi,\varphi)]\!]^{\mathscr{M}} = D(\mathscr{M}) \\ \bot & \text{otherwise} \end{cases}$$

It is not difficult to show that indeed $[\![\varphi]\!]^{\mathscr{M}} = [\![\mathrm{tr}_{\mathscr{M}}(\varphi)]\!]^{\mathscr{M}}$ holds for all $\varphi \in \mathscr{L}_{\mathsf{Kh}}$. Thus, since every $\mathscr{L}_{\mathsf{Kh}}$-definable $U \subseteq D(\mathscr{M})$ has a formula $\varphi \in \mathscr{L}_{\mathsf{Kh}}$ such that $U = [\![\varphi]\!]^{\mathscr{M}}$, there is also a propositional formula, namely $\mathrm{tr}_{\mathscr{M}}(\varphi)$, satisfying $U = [\![\mathrm{tr}_{\mathscr{M}}(\varphi)]\!]^{\mathscr{M}}$. □

Thus, given a relational model $\mathscr{M}$, a set $U \subseteq D(\mathscr{M})$ is $\mathscr{L}_{\mathsf{Kh}}$-definable if and only if it is propositionally definable. The reason for choosing propositional definability over $\mathscr{L}_{\mathsf{Kh}}$-definability is that, in this way, a bisimulation is more 'structural', defined only in terms of the model (in particular, in terms of valuations).

The second point is about the A-Zig and A-Zag clauses, which simply state that a bisimulation needs to be total for both models. As the proof of Theorem 1 will show, these conditions are required to obtain modal equivalence. Moreover, both are indeed required in the bisimulation definition, as they do not follow from Kh-Zig and Kh-Zag, although A is expressible by Kh.

Example 4. *Consider two relational models $\mathscr{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ and $\mathscr{M}' = \langle W', \mathrm{R}', V'\rangle$ as depicted below (R and R' are collections of empty relations), with a relation $Z \subseteq (W \times W')$ drawn with a dashed line.*

Clearly, $\mathscr{M}, w_1$ and $\mathscr{M}', w'_1$ are not $\mathscr{L}_{\mathsf{Kh}}$-bisimilar due to the lack of A-Zig *for w_2. However, Z clearly satisfies the* Atom *clause, and we can show that the* Kh-Zig *clause holds too by proving its four cases.*

1. *Take $U = \emptyset$. When looking for $V \subseteq W$ with $\emptyset \to V$, the only possibility is $V = \emptyset$. Since $Z[\emptyset] = \emptyset$, we need a set $V' \subseteq W'$ satisfying (i) $\emptyset \to V'$ and (ii) for each $v' \in V'$ there is $v \in \emptyset$ such that vZv'. Clearly, $V' = \emptyset$ satisfies the requirements.*

2. *Take $U = \{w_1\}$. As $\mathrm{R}_a = \emptyset$ for every $a \in \Sigma$, the only $V \subseteq W$ satisfying $\{w_1\} \to V$ is $\{w_1\}$ (via $\sigma = \varepsilon$). Since $Z[\{w_1\}] = \{w'_1\}$, we need a $V' \subseteq W'$ satisfying (i) $\{w'_1\} \to V'$ and (ii) for each $v' \in V'$ there is $v \in \{w_1\}$ such that vZv'. Clearly, $V' = \{w'_1\}$ satisfies both requirements (using $\sigma = \varepsilon$ for the first).*

3. *Take $U = \{w_2\}$. As $\mathrm{R}_a = \emptyset$ for every $a \in \Sigma$, the only $V \subseteq W$ satisfying $\{w_2\} \to V$ is $\{w_2\}$ (via $\sigma = \varepsilon$). Since $Z[\{w_2\}] = \emptyset$, we need a set $V' \subseteq W'$ satisfying (i) $\emptyset \to V'$ and (ii) for each $v' \in V'$ there is $v \in \emptyset$ such that vZv'. Clearly, $V' = \emptyset$ also satisfies both requirements.*

4. *Take $U = \{w_1, w_2\}$. As $\mathrm{R}_a = \emptyset$ for every $a \in \Sigma$, the only $V \subseteq W$ satisfying $\{w_1, w_2\} \to V$ is $\{w_1, w_2\}$ (via $\sigma = \varepsilon$). Since $Z[\{w_1, w_2\}] = \{w'_1\}$, we need a set $V' \subseteq W'$ satisfying (i) $\{w'_1\} \to V'$ and (ii) for each $v' \in V'$ there is $v \in \{w_1, w_2\}$ such that vZv'. Clearly, $V' = \{w'_1\}$ does the work.*

An analogous argument shows that the Kh*-Zag clause holds too. This shows that* Kh*-Zig and* Kh*-Zag do not imply* A*-Zig and* A*-Zag.*

Now we are ready to verify that the two models in Example 2 are indeed $\mathscr{L}_{\mathsf{Kh}}$-bisimilar.

It is also useful to note how the presented notion of bisimulation is indeed an equivalence relation among relational models, a result that is sometimes implicitly used.

Proposition 2. $\leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}}$ *is an equivalence relation.*

Proof. We will show that $\leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}}$ is reflexive, symmetric and transitive.

Reflexivity: Is trivial because the identity relation is a bisimulation.

Symmetry: Let $\mathscr{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ and $\mathscr{M}' = \langle W', \mathrm{R}', \mathrm{V}'\rangle$ be two models, $w \in W$, $w' \in W'$, such that $\mathscr{M}, w \leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}} \mathscr{M}', w'$. Then there exists a $Z \subseteq (W \times W')$ s.t. wZw'. Let us show that the relation Z^{-1} is a bisimulation.

The Atom condition is trivial. Suppose $u'Z^{-1}u$, then by definition uZu'. Let $U' \subseteq W'$ be propositional definable. By Kh-Zag, if $U' \to V'$, for some $V' \subseteq W'$, then there exists $V \subseteq W$ s.t. $Z^{-1}[U'] \to V$ and for all $v \in V$ there is $v' \in V'$ s.t. vZv', i.e., $v'Z^{-1}v$. Then Z^{-1} satisfies Kh-Zig. Similar for Kh-Zag.

Transitivity: Let $\mathscr{M}_1 = \langle W_1, \mathrm{R}_1, \mathrm{V}_1\rangle$, $\mathscr{M}_2 = \langle W_2, \mathrm{R}_2, \mathrm{V}_2\rangle$ and $\mathscr{M}_3 = \langle W_3, \mathrm{R}_3, \mathrm{V}_3\rangle$ be three models, $w_1 \in W_1$, $w_1 \in W_1$, $w_3 \in W_3$, such that $\mathscr{M}_1, w_1 \leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}} \mathscr{M}_2, w_2$ and $\mathscr{M}_2, w_2 \leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}} \mathscr{M}_3, w_3$. Then there exist Z_1 and Z_2 s.t. $w_1 Z_1 w_2$ and $w_2 Z_2 w_3$. It is easy to show that $Z_3 = Z_1 \circ Z_2$ is a $\mathscr{L}_{\mathsf{Kh}}$-bisimulation s.t. $w_1 Z_3 w_3$, then $\mathscr{M}_1, w_1 \leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}} \mathscr{M}_3, w_3$.

□

Now, for the first direction of the characterization result: bisimilarity ($\leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}}$) implies modal equivalence ($\equiv_{\mathscr{L}_{\mathsf{Kh}}}$).

Theorem 1 ($\mathscr{L}_{\mathsf{Kh}}$ Invariance)**.** *Let $\mathscr{M}, w$ and $\mathscr{M}', w'$ be two pointed models, with $\mathscr{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ and $\mathscr{M}' = \langle W', \mathrm{R}', \mathrm{V}'\rangle$. If $\mathscr{M}, w \leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}} \mathscr{M}', w'$, then $\mathscr{M}, w \equiv_{\mathscr{L}_{\mathsf{Kh}}} \mathscr{M}', w'$.*

Proof. If $\mathscr{M}, w \leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}} \mathscr{M}', w'$ then there is a $\mathscr{L}_{\mathsf{Kh}}$-bisimulation $Z \subseteq (W \times W')$ such that wZw'. The proof is by structural induction on $\mathscr{L}_{\mathsf{Kh}}$-formulas. Boolean cases are straightforward; we only prove the case for $\mathsf{Kh}(\psi, \varphi)$.

Suppose $\mathscr{M}, w \models \mathsf{Kh}(\psi, \varphi)$. Then there exists a $\sigma \in \Sigma^*$ such that $[\![\psi]\!]^{\mathscr{M}} \xrightarrow{\sigma} V$ and $\mathrm{R}_\sigma[[\![\psi]\!]^{\mathscr{M}}] \subseteq [\![\varphi]\!]^{\mathscr{M}}$. First, an auxiliary result: $Z[[\![\psi]\!]^{\mathscr{M}}] = [\![\psi]\!]^{\mathscr{M}'}$.

($\subseteq$) Take any $v' \in Z[[\![\psi]\!]^{\mathscr{M}}]$. Then there exists $v \in [\![\psi]\!]^{\mathscr{M}}$ such that vZv'. By inductive hypothesis, $v' \in [\![\psi]\!]^{\mathscr{M}'}$.

($\supseteq$) Take any $v' \in [\![\psi]\!]^{\mathscr{M}'}$. By A-Zag there exists v such that vZv'; by inductive hypothesis, $v \in [\![\psi]\!]^{\mathscr{M}}$. Then, $v' \in Z[[\![\psi]\!]^{\mathscr{M}}]$.

Now, let us show that $\mathsf{Kh}(\psi, \varphi)$ holds at $\mathscr{M}', w'$. Since $[\![\psi]\!]^{\mathscr{M}}$ is obviously $\mathscr{L}_{\mathsf{Kh}}$-definable, it is also $\mathscr{L}$-definable (Proposition 1), and hence clause Kh-Zig tells us that, for every $V \subseteq W$ such that $[\![\psi]\!]^{\mathscr{M}} \to V$, there is a $V' \subseteq W'$ such that *(i)* $Z[[\![\psi]\!]^{\mathscr{M}}] \to V'$ (i.e., $[\![\psi]\!]^{\mathscr{M}'} \to V'$, given the auxiliary result), and *(ii)* for every $v' \in V'$ there is $v \in V$ such that vZv'. From $[\![\psi]\!]^{\mathscr{M}} \to V$ and the fact that $\mathsf{Kh}(\psi, \varphi)$ holds at

$(\mathcal{M}, w)$, we know that $V \subseteq [\![\varphi]\!]^{\mathcal{M}}$. Then, from *(ii)*, for every $v' \in V'$ there is $v \in [\![\varphi]\!]^{\mathcal{M}}$ such that vZv'; hence, by inductive hypothesis, $v' \in [\![\varphi]\!]^{\mathcal{M}'}$. Thus, summarising, $[\![\psi]\!]^{\mathcal{M}'} \to V'$ and $V' \subseteq [\![\varphi]\!]^{\mathcal{M}'}$; hence, $\mathcal{M}', w' \models \mathsf{Kh}(\psi, \varphi)$.

For the other direction use Kh-Zag and A-Zig. □

Theorem 1 tells us that the semantic condition described by our bisimulation definition is strong enough to guarantee $\mathscr{L}_{\mathsf{Kh}}$-equivalence. The other direction, showing that our bisimulation is weak enough to hold between any $\mathscr{L}_{\mathsf{Kh}}$-equivalent pointed models, is more complicated in the general case. 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 [6, 5]. Here we will focus rather on *finite* models; this is because the global modality is definable in our language, and thus a finite domain is required in order to ensure the image-finiteness property.

Theorem 2. *Let $\mathcal{M} = \langle \mathrm{W}, \mathrm{R}, \mathrm{V} \rangle$, $\mathcal{M}' = \langle \mathrm{W}', \mathrm{R}', \mathrm{V}' \rangle$ be two finite models, $w \in W$ and $w' \in W'$. If $\mathcal{M}, w \equiv_{\mathscr{L}_{\mathsf{Kh}}} \mathcal{M}', w'$ then $\mathcal{M}, w \leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}} \mathcal{M}', w'$.*

Proof. Define $Z := \{(v, v') \in (W \times W') \mid \mathcal{M}, v \equiv_{\mathscr{L}_{\mathsf{Kh}}} \mathcal{M}', v'\}$. We will show that Z is a $\mathscr{L}_{\mathsf{Kh}}$-bisimulation, so take any (w, w') in Z.

Atom: Trivial, by definition of Z.

A-**Zig:** Take $v \in W$ and suppose, for the sake of a contradiction, that there is no $v' \in W'$ such that vZv'. Then, from Z's definition, for each $v'_i \in W' = \{v'_1, \ldots, v'_n\}$ (recall: $\mathcal{M}'$ is finite) there is a $\mathscr{L}_{\mathsf{Kh}}$-formula θ_i such that $\mathcal{M}, v \models \theta_i$ but $\mathcal{M}', v'_i \not\models \theta_i$. Now take $\theta := \theta_1 \wedge \cdots \wedge \theta_n$. 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 θ_i false. Then, $\mathcal{M}, w \models \mathsf{E}\theta$ and $\mathcal{M}', w' \not\models \mathsf{E}\theta$, contradicting wZw'.

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

Kh-**Zig:** Take any propositionally definable set $[\![\psi]\!]^{\mathcal{M}} \subseteq W$ (ψ is a propositional formula), and suppose $[\![\psi]\!]^{\mathcal{M}} \to V$ for some $V \subseteq W$. We need to find a $V' \subseteq W'$ such that $Z[[\![\psi]\!]^{\mathcal{M}}] \to V'$ and for all $v' \in V'$ there is $v \in V$ such that vZv'. First note that $[\![\psi]\!]^{\mathcal{M}'} = Z[[\![\psi]\!]^{\mathcal{M}}]$,[4] we just need to find an appropriate V' for $[\![\psi]\!]^{\mathcal{M}'}$. Note that if $[\![\psi]\!]^{\mathcal{M}}$ is empty then $[\![\psi]\!]^{\mathcal{M}'} = Z[[\![\psi]\!]^{\mathcal{M}}]$ is empty too and we can just let $V' = \emptyset$. In the following we assume that $[\![\psi]\!]^{\mathcal{M}}$ is not empty, then it is clear that $[\![\psi]\!]^{\mathcal{M}'} = Z[[\![\psi]\!]^{\mathcal{M}}]$ is not empty too due to A-Zag. Thus V and V' cannot be empty neither due to the definition of the forcing relation based on strong executability. Moreover, as an easy observation, there is $V' = [\![\psi]\!]^{\mathcal{M}'}$ such that $[\![\psi]\!]^{\mathcal{M}'} \to V'$ due to the empty strategy.

Now, towards a contradiction, suppose that such a V' does not exist, i.e., for each $V' \subseteq W'$ such that $[\![\psi]\!]^{\mathcal{M}'} \to V'$, there exists a $v'_{V'} \in V'$ such that there is no $v \in V$ such that $vZv'_{V'}$. Due to the definition of Z, this means that for each $v \in V$ we have a formula $\varphi^v_{V'}$ such that $\mathcal{M}, v \vDash \varphi^v_{V'}$ but $\mathcal{M}', v' \nvDash \varphi^v_{V'}$. Now since the models are finite, we can define $\theta_{V'} := \bigvee_{v \in V} \varphi^v_{V'}$, then let $\theta := \bigwedge_{\{V' \mid [\![\psi]\!]^{\mathcal{M}'} \to V'\}} \theta_{V'}$. Then it is not hard to see that $\mathcal{M}, v \vDash \theta$ for all $v \in V$ but there is a v' in V' such that $\mathcal{M}', v' \nvDash \theta$ for each V' such that $[\![\psi]\!]^{\mathcal{M}'} \to V'$. Therefore, since Kh-formulas are global, $\mathcal{M}, w \models \mathsf{Kh}(\psi, \theta)$ but $\mathcal{M}', w' \not\models \mathsf{Kh}(\psi, \theta)$. Contradiction.

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

Thus, Z is a $\mathscr{L}_{\mathsf{Kh}}$-bisimulation, and therefore $\mathcal{M}, w \leftrightarrows_{\mathscr{L}_{\mathsf{Kh}}} \mathcal{M}', w'$. □

[4] ($\supseteq$) $u' \in Z[[\![\psi]\!]^{\mathcal{M}}]$ implies there is $u \in [\![\psi]\!]^{\mathcal{M}}$ such that uZu', and therefore, from Z's definition, $u' \in [\![\psi]\!]^{\mathcal{M}'}$. ($\subseteq$) Suppose $u' \in [\![\psi]\!]^{\mathcal{M}'}$. From A-Zag there is $u \in W$ such that uZu'; then, from Z's definition, $u \in [\![\psi]\!]^{\mathcal{M}}$ and therefore $u' \in Z[[\![\psi]\!]^{\mathcal{M}}]$.

Note that the statement of Theorem 2 only indicates that *finite* models that are $\mathscr{L}_{\mathsf{Kh}}$-equivalent are also bisimilar. Note also that, as the language can express the universal modality, merely restricting to image-finite models as in standard modal logic does not work.

4 Knowing How with Intermediate Constraints

The knowing-how operator Kh has a 'conditional' feeling: $\mathcal{M}, w \models \mathsf{Kh}(\psi, \varphi)$ holds if and only if the agent has a 'method' σ whose output is guaranteed to be φ, whenever ψ 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 ψ-worlds to φ-worlds, any $\sigma \in \Sigma^*$ will do.

Of course, 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 certain condition. This is the idea behind the *knowing how* operator with intermediate constraints studied in [18]. This section introduces a notion of bisimulation for a logic based on this operator, Khm.

Definition 8 (Syntax)**.** *We define the set of $\mathscr{L}_{\mathsf{Khm}}$-formulas as:*

$$\varphi ::= \bot \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{Khm}(\varphi, \varphi, \varphi),$$

with $p \in \mathsf{PROP}$. Formulas of the form $\mathsf{Khm}(\psi, \chi, \varphi)$ express that, given ψ, the agent knows how to achieve φ while maintaining χ. Additionally, the abbreviation $\mathsf{A}\varphi$ is given now as $\mathsf{A}\varphi := \mathsf{Khm}(\neg\varphi, \top, \bot)$ (and $\mathsf{E}\varphi := \neg\mathsf{A}\neg\varphi$, as before).

For Khm's semantic interpretation, some further definitions are required.

Definition 9. *Let $\mathcal{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ be a relational model.*

- *We say that $\sigma = a_1 \dots a_n$ is* strongly χ-executable at *$w \in W$ if and only if (i) σ is strongly executable at w, and (ii) for all $0 < k < n$, if $v \xrightarrow{\sigma_k} t$ then $\mathcal{M}, t \models \chi$.*
- *Let $\sigma = a_1 \dots a_n$ a sequence in Σ^*. Given $X \subseteq W$ and $w \in W$, we say σ is* strongly X-executable at *w if and only if (i) σ is strongly executable at w, and (ii) $\{v \mid w \xrightarrow{\sigma_k} v$, for any $0 < k < n\} \subseteq X$.*
- *We write $U \xrightarrow{\sigma, X} V$ if and only if (i) σ is strongly X-executable at every $u \in U$, and (ii) $V = \mathrm{R}_\sigma[U]$.*
- *We write $U \xrightarrow{X} V$ whenever there is a $\sigma \in \Sigma^*$ such that $U \xrightarrow{\sigma, X} V$.*

Now we can introduce the semantics of the Khm operator.

Definition 10 (Semantics)**.** *Let $\mathcal{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ be a model, and take $w \in W$. The satisfaction relation $\models$ for atoms, negations and conjunctions is as in Definition 4. For Khm,*

$\mathcal{M}, w \models \mathsf{Khm}(\psi, \chi, \varphi)$	*iff*	*there exists $\sigma \in \Sigma^*$ such that for all $u \in W$, if $\mathcal{M}, u \models \psi$, then σ is strongly χ-executable at u and, for all $v \in W$, $u \xrightarrow{\sigma} v$ implies $\mathcal{M}, v \models \varphi$.*

Note how, given its new definition (Definition 8), A is again the global universal modality.

Here is, then, a bisimulation for $\mathscr{L}_{\mathsf{Khm}}$.

Definition 11. *Let $\mathcal{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ and $\mathcal{M}' = \langle W', \mathrm{R}', \mathrm{V}'\rangle$ be two relational models. A non-empty relation $Z \subseteq (W \times W')$ is called an $\mathscr{L}_{\mathsf{Khm}}$-bisimulation between $\mathcal{M}$ and $\mathcal{M}'$ if and only if wZw' implies:*

Atom, A**-Zig and** A**-Zag** *as in Definition 7.*

Khm**-Zig:** *for any propositional definable $U \subseteq W$, if $U \xrightarrow{X} V$ for some $X, V \subseteq W$, then there are $X', V' \subseteq W'$ such that (i) $Z[U] \xrightarrow{X'} V'$, (ii) for each $x' \in X'$ there is a $x \in X$ such that xZx', and (iii) for each $v' \in V'$ there is a $v \in V$ such that vZv'.*

Khm**-Zag:** *for any propositional definable $U' \subseteq W'$, if $U' \xrightarrow{X'} V'$ for some $X', V' \subseteq W'$, then there are $X, V \subseteq W$ such that (i) $Z^{-1}[U'] \xrightarrow{X} V$, (ii) for each $x \in X$ there is a $x' \in X'$ such that xZx', and (iii) for each $v \in V$ there is a $v' \in V'$ such that vZv'.*

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

The difference between clauses Kh-Zig/Zag and Khm-Zig/Zag is that the latter also require for the 'travelled states' to be bisimilar. It is also worthwhile to notice that the *propositional definability* requirement is kept since, just as with $\mathcal{L}_{\mathsf{Kh}}$, a set $U \subseteq D(\mathcal{M})$ is $\mathcal{L}_{\mathsf{Khm}}$-definable if and only if it is propositionally definable. We will use the notion of Khm-equivalence ($\equiv_{\mathcal{L}_{\mathsf{Khm}}}$) as in Definition 5 but for Khm-formulas.

Now we will show that $\mathcal{L}_{\mathsf{Khm}}$-bisimilarity implies $\mathcal{L}_{\mathsf{Khm}}$-equivalence.

Theorem 3 ($\mathcal{L}_{\mathsf{Khm}}$ Invariance)**.** *Let $\mathcal{M}, w$ and $\mathcal{M}', w'$ be two pointed models, with $\mathcal{M} = \langle W, R, V \rangle$ and $\mathcal{M}' = \langle W', R', V' \rangle$. If $\mathcal{M}, w \leftrightarroweq_{\mathcal{L}_{\mathsf{Khm}}} \mathcal{M}', w'$, then $\mathcal{M}, w \equiv_{\mathcal{L}_{\mathsf{Khm}}} \mathcal{M}', w'$.*

Proof. If $\mathcal{M}, w \leftrightarroweq_{\mathcal{L}_{\mathsf{Khm}}} \mathcal{M}', w'$ then there is a $\mathcal{L}_{\mathsf{Khm}}$-bisimulation $Z \subseteq (W \times W')$ such that wZw'. The proof is by structural induction on $\mathcal{L}_{\mathsf{Khm}}$-formulas. Boolean cases are straightforward; we only prove the case for $\mathsf{Khm}(\psi, \chi, \varphi)$.

Suppose $\mathcal{M}, w \models \mathsf{Khm}(\psi, \chi, \varphi)$. Then there is $\sigma \in \Sigma^*$ such that $[\![\psi]\!]^{\mathcal{M}} \xrightarrow{\sigma, X} V$ (for $X \subseteq [\![\chi]\!]^{\mathcal{M}}$) and $V \subseteq [\![\varphi]\!]^{\mathcal{M}}$. The set $[\![\psi]\!]^{\mathcal{M}}$ is $\mathcal{L}_{\mathsf{Khm}}$-definable, and thus $\mathcal{L}$-definable. Hence, from $[\![\psi]\!]^{\mathcal{M}} \xrightarrow{\sigma, X} V$, the Khm-Zig clause tells us that there are $X', V' \subseteq W'$ such that *(i)* $Z[[\![\psi]\!]^{\mathcal{M}}] \xrightarrow{X'} V'$, *(ii)* for each $v' \in V'$ there is a $v \in V$ such that vZv', and *(iii)* for each $x' \in X'$ there is a $x \in X$ such that xZx'.

We now prove three parts. First, from *(i)* and $Z[[\![\psi]\!]^{\mathcal{M}}] = [\![\psi]\!]^{\mathcal{M}'}$ (see the proof of Theorem 1) we have $[\![\psi]\!]^{\mathcal{M}'} \xrightarrow{X'} V'$. Second, if $x' \in X'$, then from *(iii)* there is $x \in X$ such that xZx'. Since $X \subseteq [\![\chi]\!]^{\mathcal{M}}$, there is $x \in [\![\chi]\!]^{\mathcal{M}}$ such that xZx'; therefore, by induction hypothesis, $x' \in [\![\chi]\!]^{\mathcal{M}'}$. Hence, $X' \subseteq [\![\chi]\!]^{\mathcal{M}'}$. Third, take any $v' \in V'$. From *(ii)*, there is $v \in V$ such that vZv'; but $V \subseteq [\![\varphi]\!]^{\mathcal{M}}$, so $v \in [\![\varphi]\!]^{\mathcal{M}}$ and vZv'. Then, by induction hypothesis, $v' \in [\![\varphi]\!]^{\mathcal{M}'}$, so $V' \subseteq [\![\varphi]\!]^{\mathcal{M}'}$. Thus, $[\![\psi]\!]^{\mathcal{M}'} \xrightarrow{X'} V'$, $X' \subseteq [\![\chi]\!]^{\mathcal{M}'}$ and $V' \subseteq [\![\varphi]\!]^{\mathcal{M}'}$; therefore, $\mathcal{M}', w' \models \mathsf{Khm}(\psi, \chi, \varphi)$. The argument for the other direction is similar. □

For the other direction, in *finite* models, $\mathcal{L}_{\mathsf{Khm}}$-equivalence implies $\mathcal{L}_{\mathsf{Khm}}$-bisimilarity.

Theorem 4. *Let $\mathcal{M} = \langle W, R, V \rangle$ and $\mathcal{M}' = \langle W', R', V' \rangle$ be two finite models, $w \in W$ and $w' \in W'$. If $\mathcal{M}, w \equiv_{\mathcal{L}_{\mathsf{Khm}}} \mathcal{M}', w'$ then $\mathcal{M}, w \leftrightarroweq_{\mathcal{L}_{\mathsf{Khm}}} \mathcal{M}', w'$.*

Proof. The strategy will be, just as in the proof of Theorem 2, to show that $\mathcal{L}_{\mathsf{Khm}}$-equivalence is indeed a bisimulation, so define $Z := \{(v, v) \in (W \times W') \mid \mathcal{M}, v \equiv_{\mathcal{L}_{\mathsf{Khm}}} \mathcal{M}', v'\}$ and take any (w, w') in Z.

Atom: Trivial, by definition of Z. A-Zig and A-Zag are as in Theorem 2.

Khm**-Zig:** Take any propositionally definable $[\![\psi]\!]^{\mathcal{M}} \subseteq W$, and suppose $[\![\psi]\!]^{\mathcal{M}} \xrightarrow{X} V$ for some $X, V \subseteq W$. As in the proof of Theorem 2, we can show that $[\![\psi]\!]^{\mathcal{M}'} = Z[[\![\psi]\!]^{\mathcal{M}}]$. According to the definition of Khm-Zig, we just need to show that $[\![\psi]\!]^{\mathcal{M}'} \xrightarrow{X'} V'$ for some $X', V' \subseteq W'$ such that (ii) for each $x' \in X'$ there is an $x \in X$ such that xZx', and (iii) for each $v' \in V'$ there is a $v \in V$ such that vZv'.

As a handy observation, we can show that $[\![\psi]\!]^{\mathcal{M}'} \xrightarrow{X'} V'$ for some appropriate X', V' satisfying (ii) and (iii) iff $[\![\psi]\!]^{\mathcal{M}'} \xrightarrow{Z[X]} V'$ for some appropriate V' satisfying (iii). Note that right-to-left is trivial since $Z[X]$ satisfies (ii) by definition. Left-to-right is also clear since $X' \subseteq Z[X]$ if X' satisfies (ii). By the above observation, we just need to show that $[\![\psi]\!]^{\mathcal{M}'} \xrightarrow{Z[X]} V'$ for some appropriate V' satisfying (iii).

Now since the models are finite, we can find a propositional formula χ to characterize $Z[X]$ within $\mathcal{M}'$. To see this, note that for each valuation for basic propositions, we can use a finite propositional formula to distinguish it from finitely many other valuations. Moreover, if two worlds u', u'' share the same valuation then they are either both in $Z[X]$ or both outside $Z[X]$, for otherwise there exists $x \in X$ such that xZu' but not xZu'', which is impossible since u and u'' are logically equivalent (the global Kh-formulas do not distinguish them).

The rest of the proof is very similar to the one for Theorem 2. Towards a contradiction, suppose that any V' such that $[\![\psi]\!]^{\mathcal{M}'} \xrightarrow{Z[X]} V'$ does not satisfy (iii) for the given V. Then we can find the θ in a similar way as in the proof of Theorem 2. Finally we can show that $\mathcal{M}, w \models \mathsf{Khm}(\psi, \chi, \theta)$ but $\mathcal{M}', w' \not\models \mathsf{Khm}(\psi, \chi, \theta)$, where χ is a formula which characterizes $Z[X]$ within $\mathcal{M}'$. Contradiction.

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

Thus, Z is a $\mathscr{L}_{\mathsf{Khm}}$-bisimulation, and therefore $\mathcal{M}, w \,\underline{\leftrightarrow}_{\mathscr{L}_{\mathsf{Khm}}} \mathcal{M}', w'$. □

5 A Weaker Logic of Knowing How

Sometimes, the logics of *knowing how* introduced in previous chapters are too strong, given that they require plans to be strongly executable. In [17] a *knowing how* operator based on weak conformant plans was introduced. A weak conformant plan for achieving φ-states from ψ-states is a finite linear action sequence such that the execution of the action sequence at each ψ-state will always terminate on a φ-state, either successfully or not. The motivation for introducing a weaker operator is that for real life situations a weak conformant plan is enough. For instance, drinking ten shots of tequila sounds like a plan for getting drunk, but the weaker conformant plan consisting of drinking nine (and even less) shots and stop may achieve the goal.

Definition 12 (Syntax)**.** *We define the set of $\mathscr{L}_{\mathsf{Kh^W}}$-formulas as:*

$$\varphi ::= \bot \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{Kh^W}(\varphi, \varphi),$$

with $p \in \mathsf{PROP}$. Formulas of the form $\mathsf{Kh^W}(\psi, \varphi)$ express that the agent can always execute a plan from ψ-states which terminates in a φ-state, either successfully or not. Additionally, the abbreviation $\mathsf{A}\varphi$ is now defined as $\mathsf{A}\varphi := \mathsf{Kh^W}(\neg\varphi, \bot)$ (and $\mathsf{E}\varphi := \neg\mathsf{A}\neg\varphi$, as before).

For $\mathsf{Kh^W}$'s semantic interpretation, some further definitions are required.

Definition 13. *Let $\mathcal{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ be a relational model.*

- *Let $\sigma = a_1 \ldots a_n$ a sequence in Σ^*, $w \in W$. $\mathsf{Terms}(w, \sigma)$ is the set of states at which executing σ on w might terminate. Formally, it is defined as*

$$\mathsf{Terms}(w, \sigma) = \{v \mid w \xrightarrow{\sigma} v, \text{or there exists } i \text{ such that } w \xrightarrow{\sigma_i} v \text{ and } v \text{ has no } a_{i+1} \text{ successor}\}.$$

- *We write $U \xrightarrow{\sigma}_{\mathsf{W}} V$ whenever $V = \bigcup_{u_i \in U} \mathsf{Terms}(u_i, \sigma)$.*

- *We write $U \to_{\mathsf{W}} V$ whenever there is σ such that $U \xrightarrow{\sigma}_{\mathsf{W}} V$.*

Now let us introduce the semantics of Kh^{W}:

Definition 14. *Let $\mathcal{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ be a model, and take $w \in W$. The satisfaction relation $\models$ for atoms, negations and conjunctions is as in Definition 4. For Kh^{W},*

$$\mathcal{M}, w \models \mathsf{Kh}^{\mathsf{W}}(\psi, \varphi) \quad \textit{iff} \quad \text{there exists } \sigma \in \Sigma^* \text{ such that for all } v \in W, \text{ if } \mathcal{M}, v \models \psi \text{ then for all } t \in \mathsf{Terms}(v, \sigma),, \text{ we have } \mathcal{M}, t \models \varphi.$$

Notice that again, the universal operator A *has the intended meaning.*

Here we have a notion of bisimulation for $\mathcal{L}_{\mathsf{Kh}^{\mathsf{W}}}$.

Definition 15. *A non-empty relation $Z \subseteq (W \times W')$ is called an $\mathcal{L}_{\mathsf{Kh}^{\mathsf{W}}}$-bisimulation if wZw' implies:*

Atom, A-Zig and A-Zag *as in Definition 7.*

Khm-**Zig** *for any propositional definable $U \subseteq W$, if $U \to_{\mathsf{W}} V$ for some $V \subseteq W$, then (i) there is $V' \subseteq W'$ such that $\mathrm{Z}[U] \to_{\mathsf{W}} V'$ and, (ii) for each $v' \in V'$ there is a $v \in V$ such that vZv'.*

Khm-**Zag** *for any propositional definable $U' \subseteq W'$, if (i) $U' \to_{\mathsf{W}} V'$ for some $V' \subseteq W'$, then there is $V \subseteq W$ such that $\mathrm{Z}^{-1}[U'] \to_{\mathsf{W}} V$ and, (ii) for each $v \in V$ there is a $v' \in V'$ such that vZv'.*

We write $\mathcal{M}, w \leftrightarroweq_{\mathcal{L}_{\mathsf{Kh}^{\mathsf{W}}}} \mathcal{M}', w'$ when there is a $\mathcal{L}_{\mathsf{Kh}^{\mathsf{W}}}$-bisimulation Z such that wZw'.

In a very similar way as we did in previous sections, we can prove the following two theorems (using, of course, the appropriate definition of $\equiv_{\mathcal{L}_{\mathsf{Kh}^{\mathsf{W}}}}$).

Theorem 5 **($\mathcal{L}_{\mathsf{Kh}}$ Invariance).** *Let $\mathcal{M}, w$ and $\mathcal{M}', w'$ be two pointed models, with $\mathcal{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$ and $\mathcal{M}' = \langle W', \mathrm{R}', \mathrm{V}'\rangle$. If $\mathcal{M}, w \leftrightarroweq_{\mathcal{L}_{\mathsf{Kh}^{\mathsf{W}}}} \mathcal{M}', w'$, then $\mathcal{M}, w \equiv_{\mathcal{L}_{\mathsf{Kh}^{\mathsf{W}}}} \mathcal{M}', w'$.*

Theorem 6. *Let $\mathcal{M} = \langle W, \mathrm{R}, \mathrm{V}\rangle$, $\mathcal{M}' = \langle W', \mathrm{R}', \mathrm{V}'\rangle$ be two finite models, $w \in W$ and $w' \in W'$. If $\mathcal{M}, w \equiv_{\mathcal{L}_{\mathsf{Kh}^{\mathsf{W}}}} \mathcal{M}', w'$ then $\mathcal{M}, w \leftrightarroweq_{\mathcal{L}_{\mathsf{Kh}^{\mathsf{W}}}} \mathcal{M}', w'$.*

6 Final Remarks

In this article we introduced suitable notions of bisimulation for three different *knowing how* logics. These logics represent an interesting alternative for a new interpretation of epistemic logics, traditionally focused on *knowing that* operations. The semantics of the first *knowing how* operator (first introduced in [23, 25]) requires plans that always succeed in order to know how to achieve φ from a precondition ψ. The second alternative (introduced in [18]) incorporates conditions in each intermediate state that is "visited" while executing the successful plan. Finally, a weaker *knowing how* operator was introduced in [17], in which it is considered that the goal φ 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. As we pointed out, these operators were already investigated but the model theoretical aspects have not been studied before. For the three notions of bisimulation introduced, we proved that bisimilarity implies modal equivalence in the corresponding logic, and when consider the class of finite models, modal equivalence implies bisimilarity.

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, we will use bisimulations as a tool to investigate the expressive power of each logic, and compare them among each other. Also, bisimulations can be used to minimize models as mentioned before. It would be interesting to explore also model theoretical properties for the logics which can capture interactions between *knowing how* and *knowing that* operators (such as the logics introduced in [9, 20]). Of course, another interesting extension is to move to the multi-agent setting, in order to describe the *knowing how* abilities of different agents. When combined with a *knowledge 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.

Acknowledgements

Raul Fervari has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No. 690974 for the project MIREL: MIning and REasoning with Legal texts, SeCyT-UNC and ANPCyT-PICT-2013-2011. Yanjing Wang acknowledges the support from the National Program for Special Support of Eminent Professionals, and the NSSF key project 12&ZD119.

References

[1] Thomas Ågotnes, Valentin Goranko & Wojciech Jamroga (2007): *Alternating-time temporal logics with irrevocable strategies*. In Dov Samet, editor: *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2007), Brussels, Belgium, June 25-27, 2007*, pp. 15–24, doi:10.1145/1324249.1324256. Available at `http://doi.acm.org/10.1145/1324249.1324256`.

[2] Thomas Ågotnes, Valentin Goranko, Wojciech Jamroga & Michael Wooldridge (2015): *Knowledge and Ability*. In van Ditmarsch et al. [7], chapter 11, pp. 543–589.

[3] Carlos Areces & Diego Figueira (2009): *Which Semantics for Neighbourhood Semantics?* In Craig Boutilier, editor: *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009*, pp. 671–676. Available at `http://ijcai.org/Proceedings/09/Papers/117.pdf`.

[4] Johan van van Benthem (2014): *Logic in Games*. The MIT Press.

[5] Patrick Blackburn & Johan van Benthem (2006): *Modal Logic: A Semantic Perspective*. In: *Handbook of Modal Logic*, Elsevier, pp. 1–84.

[6] Patrick Blackburn, Maarten de Rijke & Yde Venema (2002): *Modal Logic*. Cambridge University Press.

[7] Hans van Ditmarsch, Joseph Y. Halpern, Wiebe van der Hoek & Barteld Kooi, editors (2015): *Handbook of Epistemic Logic*. College Publications.

[8] Ronald Fagin, Joseph Y. Halpern, Yoram Moses & Moshe Y. Vardi (1995): *Reasoning about knowledge*. The MIT Press, Cambridge, Mass.

[9] Raul Fervari, Andreas Herzig, Yanjun Li & Yanjing Wang (2017): *Strategically knowing how*. In: *Proceedings of IJCAI 17*. To appear.

[10] Paul Gochet (2013): *An Open Problem in the Logic of Knowing How*. In Jaakko Hintikka, editor: *Open Problems in Epistemology*, The Philosophical Society of Finland.

[11] Helle Hvid Hansen (2003): *Monotonic Modal Logics*. Master's thesis, Universiteit van Amsterdam.

[12] Helle Hvid Hansen, Clemens Kupke & Eric Pacuit (2007): *Bisimulation for Neighbourhood Structures*. In Till Mossakowski, Ugo Montanari & Magne Haveraaen, editors: *Algebra and Coalgebra in Computer Science, Second International Conference, CALCO 2007, Bergen, Norway, August 20-24, 2007, Proceedings*, *Lecture Notes in Computer Science* 4624, Springer, pp. 279–293, doi:10.1007/978-3-540-73859-6_19.

[13] Helle Hvid Hansen, Clemens Kupke & Eric Pacuit (2009): *Neighbourhood Structures: Bisimilarity and Basic Model Theory*. *Logical Methods in Computer Science* 5(2), doi:10.2168/LMCS-5(2:2)2009. Available at `http://arxiv.org/abs/0901.4430`.

[14] Andreas Herzig (2015): *Logics of knowledge and action: critical analysis and challenges*. *Autonomous Agents and Multi-Agent Systems* 29(5), pp. 719–753, doi:10.1007/s10458-014-9267-z.

[15] Jaakko Hintikka (1962): *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca N.Y.

[16] Wojciech Jamroga & Thomas Ågotnes (2007): *Constructive knowledge: what agents can achieve under imperfect information*. *Journal of Applied Non-Classical Logics* 17(4), pp. 423–475, doi:10.3166/jancl.17.423-475.

[17] Yanjun Li (2017): *Stopping Means Achieving: A Weaker Logic of Knowing How*. *Studies in Logic* 9(4), pp. 34–54.

[18] Yanjun Li & Yanjing Wang (2017): *Achieving While Maintaining: - A Logic of Knowing How with Intermediate Constraints*. In Sujata Ghosh & Sanjiva Prasad, editors: *Logic and Its Applications - 7th Indian Conference, ICLA 2017, Kanpur, India, January 5-7, 2017, Proceedings*, *Lecture Notes in Computer Science* 10119, Springer, pp. 154–167, doi:10.1007/978-3-662-54069-5_12.

[19] John-Jules Ch. Meyer & Wiebe van Der Hoek (1995): *Epistemic Logic for AI and Computer Science*. Cambridge University Press, New York, N.Y., U.S.A.

[20] Pavel Naumov & Jia Tao: *Together We Know How to Achieve: An Epistemic Logic of Know-How*. In: *Proceedings of TARK'17*. Available at `arxiv.org/abs/1705.09349`.

[21] M. Pauly (1999): *Bisimulation for general non-normal modal logic*. Unpublished manuscript.

[22] Jason Stanley & Timothy Williamson (2001): *Knowing how*. *The Journal of Philosophy*, pp. 411–444.

[23] Yanjing Wang (2015): *A Logic of Knowing How*. In: *Proceedings of LORI 2015*, pp. 392–405.

[24] Yanjing Wang (2016): *Beyond knowing that: a new generation of epistemic logics*. In Hans van Ditmarsch & Gabriel Sandu, editors: *Jaakko Hintikka on knowledge and game theoretical semantics*, Springer. Forthcoming.

[25] Yanjing Wang (2016): *A logic of goal-directed knowing how*. *Synthese*, pp. 1–21. In press.

[26] Yanjing Wang (2017): *A new modal framework for epistemic logic*. In: *Proceedings of TARK '17*.

[27] Yanjing Wang & Yanjun Li (2012): *Not All Those Who Wander Are Lost: Dynamic Epistemic Reasoning in Navigation*. In: *Proceedings of AiML 2012*, pp. 559–580.