

On the definability of simulability and bisimilarity by finite epistemic models

Hans van Ditmarsch, David Fernandez-Duque, and Wiebe van der Hoek

University of Sevilla, Spain, {hvd,dfduque}@us.es

University of Liverpool, United Kingdom, Wiebe.Van-Der-Hoek@liverpool.ac.uk

Abstract. We explore when finite epistemic models are definable up to simulability or bisimulation, either over the basic multi-agent epistemic language L or over its extension L^C with common knowledge operators. Our negative results are that: simulability is not definable in general in L^C , and finite epistemic states (i.e., pointed models) are not definable up to bisimulation in L . Our positive results are that: finite epistemic states are definable up to bisimulation by *model validity* of L -formulas, and there is a class of epistemic models we call *well-multifounded* for which simulability is definable over L . From our method it also follows that finite epistemic models (i.e., not-pointed models) are definable up to bisimulation using model validity in L . Our results may prove useful for the logical specification of multi-agent systems, as it provides justification for the ubiquitous but often unjustified claims of the form ‘suppose action a can only be performed in state s ’: we show when such preconditions exist. An application are characteristic formulae for interpreted systems. They have a special form wherein factual knowledge, positive knowledge, and ignorance can be separated.

1 Introduction

Modal logic is *the* framework for formalising knowledge representation and areas in artificial intelligence as diverse as distributed computing, reasoning about programs, verifying temporal properties of systems, game theoretic reasoning, reasoning about knowledge and belief, and specifying and verifying multi-agent systems. One of the reasons of the popularity of modal logic in such diverse fields is its semantics: the notion of *states*, or *worlds*, together with that of a relation between them, is the key concept in Kripke models, on which modal logics are interpreted. Such worlds may model the state of a distributed system, a processor, or a machine, or a situation in a game or a protocol, and the binary relations indicate for instance a possible transition (in time, or by a computation) between states, or they may represent some attitude of an agent: some state may be desired by an agent, some states may form the goal of an agent, or, as is the interpretation for epistemic logic, states may be conceived as indistinguishable by the agent.

Kripke models are a great tool for designing and modelling complex situations, and a modal language provides a perfect way to verify properties about the

systems obtained. To give an example, consider the simplest multi-agent system one can envisage to reason about information: we assume to have two agents, a and b , and one atomic fact, say p . Suppose the information about the scenario to be modelled is the following: p is true, although neither a nor b knows it. One model that comes to mind to represent this situation is \mathcal{M}_1 depicted in Figure 1, the ‘actual state’ is s_1 . Given this state where p is true, each agent considers another state possible were it is false.

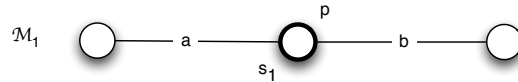


Fig. 1. A two-agent one-fact scenario.

However, this raises many questions: is this *the* model of the scenario, or are there alternative models, and how do we tell they are different? For instance, would the models of Figure 2 be ‘equally good’? It of course depends on what is meant by that criterion: yes, all three models satisfy the description of the scenario ($p \wedge \neg K_a p \wedge \neg K_b p$ is true in all of their designated states s_i), yet each model verifies some additional and *different* information about which the scenario description stayed ambiguous, or, rather, under-specified.

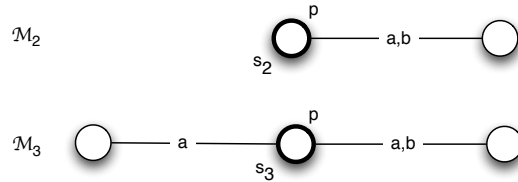


Fig. 2. Two other two-agent one-fact scenarios.

The notion of *bisimulation* [3] tells us when two structures are essentially the same, at least as far as modal logic is concerned; in this sense it plays a role analogous to, say, an isomorphism between algebraic structures. Following this analogy a *simulation* [1] would play the role of a homomorphism; simulations are defined like bisimulations without the ‘back’ clause, and as in the case of isomorphisms, bisimulations can be defined as simulations whose inverses are also simulations. These relations preserve a fragment of the modal language known as the *positive existential formulas* [3].

Referring back to our examples, we want to specify that we are actually in model \mathcal{M}_2 , in state s_2 : Can we specify that in our object language? Because the modal language is bisimulation-invariant we can hope at most to describe models up to bisimulation. However, it is not immediate that this can be done and, in fact, it is usually not the case. For example, finite unimodal K-states (i.e., arbitrary finite Kripke frames) cannot be defined up to bisimulation over the basic modal language, but they can over PDL [9, 2].

Over the class of transitive models we do have that finite pointed models are definable up to bisimulation in the basic modal language, but greater expressive power is needed to define simulability [6]. Over S5 we should expect the situation to become much simpler because accessibility is an equivalence relation [8] – indeed, both simulability by and bisimilarity to finite models are modally definable in this class – but once we consider multiagent models, things become trickier.

Our goal is, for S5, to explore when finite epistemic models are definable up to simulability or bisimulation, either over the basic modal language \mathbf{L} or the language \mathbf{L}^C enriched with common knowledge operators. Our main results are

1. simulability is not definable in general in \mathbf{L}^C (Theorem 2);
2. finite epistemic states are not definable up to bisimulation in \mathbf{L} (Theorem 1);
3. finite epistemic models are defined up to bisimulation by *model validity* in \mathbf{L} (Theorem 3), and
4. there is a class of epistemic models we call *well-multifounded* for which simulability is definable over \mathbf{L} (Theorem 5).

Note that finite epistemic states are definable up to bisimulation in \mathbf{L}^C [10]; our third result is a variant of this, exploiting the fact that the common knowledge operator is used in a somewhat shallow fashion there.

2 Epistemic logic

We consider the basic language of epistemic logic $\mathbf{L} = \mathbf{L}^A$, where A is a non-empty finite set of ‘agents’ (in this paper mainly a and b) and whose formulas are built from propositional variables in a *finite* set PV using the Boolean connectives \wedge and \neg (all other connectives are to be defined in terms of these) and the unary modal operator K_a for each $a \in A$. We write M_a as a shorthand for $\neg K_a \neg$. The reason we work with a finite set PV of propositional variables is that, evidently, one cannot define models up to simulation or bisimulation in the presence of infinitely many variables using a finite formula.

The language \mathbf{L}^C is an extension of \mathbf{L} which introduces an operator C_B (‘common knowledge’) for each $B \subseteq A$.

We are interested in interpreting \mathbf{L} and \mathbf{L}^C over *epistemic models*, which are tuples

$$\mathcal{M} = \langle |\mathcal{M}|, \sim_{\mathcal{M}}, \llbracket \cdot \rrbracket_{\mathcal{M}} \rangle$$

where $|\mathcal{M}|$ is a non-empty set, $\sim_{\mathcal{M}}$ a tuple of equivalence relations $\sim_a^{\mathcal{M}}$ for each $a \in A$ and

$$\llbracket \cdot \rrbracket_{\mathcal{M}} : PV \rightarrow 2^{|\mathcal{M}|}.$$

We will usually omit the subindex on $\sim_{\mathcal{M}}$ unless this may lead to confusion. The valuation $\llbracket \cdot \rrbracket_{\mathcal{M}}$ is extended to arbitrary formulas in the standard way for Booleans. For the epistemic modal operator we have

$$\llbracket K_a \varphi \rrbracket_{\mathcal{M}} = \{s \in |\mathcal{M}| : \forall t \stackrel{a}{\sim} s, t \in \llbracket \varphi \rrbracket_{\mathcal{M}}\},$$

and $\llbracket C_B \varphi \rrbracket_{\mathcal{M}}$ is the largest subset F of $|\mathcal{M}|$ such that, if $s \in F$, $b \in B$ and $t \stackrel{b}{\sim} s$, then $t \in \llbracket \varphi \rrbracket_{\mathcal{M}}$. For $t \in \llbracket \varphi \rrbracket_{\mathcal{M}}$ we also write $\langle \mathcal{M}, t \rangle \models \varphi$, and if $\langle \mathcal{M}, t \rangle \models \varphi$ for all $t \in |\mathcal{M}|$ we write $\mathcal{M} \models \varphi$, and we say φ is valid on model \mathcal{M} . An *epistemic state* (or *pointed epistemic model*) is a pair $\langle \mathcal{M}, s \rangle$ where \mathcal{M} is an epistemic model and $s \in |\mathcal{M}|$.

3 Simulation and bisimulation

In this section we define *simulations* and *bisimulations*, and some notions and results linking these to the logical language.

Definition 1. *If \mathcal{M}, \mathcal{N} are epistemic models, a simulation between \mathcal{M} and \mathcal{N} is a binary relation*

$$S \subseteq |\mathcal{M}| \times |\mathcal{N}| \text{ such that}$$

Atoms for every $x S y$ and every propositional variable p , $x \in \llbracket p \rrbracket_{\mathcal{M}} \Leftrightarrow y \in \llbracket p \rrbracket_{\mathcal{N}}$ and

Forth if $x' \stackrel{a}{\sim}_{\mathcal{M}} x$ and $x S y$, there is $y' \stackrel{a}{\sim}_{\mathcal{N}} y$ with $x' S y'$.

The relation S is a bisimulation if it further satisfies

Back if $x S y$ and $y' \stackrel{a}{\sim}_{\mathcal{N}} y$, there is $x' \stackrel{a}{\sim}_{\mathcal{M}} x$ with $x' S y'$.

Given models \mathcal{M} and \mathcal{N} , a point $x \in |\mathcal{M}|$ *simulates* $y \in |\mathcal{N}|$ if there exists a simulation $S \subseteq |\mathcal{M}| \times |\mathcal{N}|$ such that $x S y$; we will write $\langle \mathcal{M}, x \rangle \sqsubseteq \langle \mathcal{N}, y \rangle$. We also say that $\langle \mathcal{M}, x \rangle$ simulates $\langle \mathcal{N}, y \rangle$, or that $\langle \mathcal{N}, y \rangle$ is simulated by $\langle \mathcal{M}, x \rangle$.

If a bisimulation B exists between \mathcal{M} and \mathcal{N} such that $x B y$, we will write $\langle \mathcal{M}, x \rangle \Leftrightarrow \langle \mathcal{N}, y \rangle$ or (unless confusion results) $x \Leftrightarrow y$. We write $\mathcal{M} \Leftrightarrow \mathcal{N}$ (in words, \mathcal{M} and \mathcal{N} are *bisimilar* to each other) if there is a bisimulation between them with domain $|\mathcal{M}|$ and range $|\mathcal{N}|$. This is called a *total bisimulation*.

A well-known fact [3] is that:

Proposition 1. *If $\langle \mathcal{M}, x \rangle \Leftrightarrow \langle \mathcal{N}, y \rangle$, then x and y satisfy the same formulas of \mathcal{L} .*

Although we are mainly interested in “full” simulations and bisimulations, we occasionally need an approximation to a bisimulation given by an *k-bisimulation*, defined as follows:

Definition 2. *Let \mathcal{M}, \mathcal{N} be epistemic models, $x \in |\mathcal{M}|$ and $y \in |\mathcal{N}|$ and $k \geq 0$. We define $x \Leftrightarrow_k y$ (x is k -bisimilar to y) if x and y satisfy the same set of atoms and either $k = 0$ or the following variant of the ‘back and forth’ conditions holds:*

Forth' if $x' \overset{a}{\sim}_{\mathcal{M}} x$, there is $y' \overset{a}{\sim}_{\mathcal{N}} y$ with $x' \rightleftharpoons_{k-1} y'$ and
Back' if $y' \overset{a}{\sim}_{\mathcal{N}} y$, there is $x' \overset{a}{\sim}_{\mathcal{M}} x$ with $x' \rightleftharpoons_{k-1} y'$.

Then we have the following, also found in [3]:

Proposition 2. *If $\langle \mathcal{M}, x \rangle \rightleftharpoons_k \langle \mathcal{N}, y \rangle$, then x and y satisfy the same formulas of L of modal depth at most k .*

Recall that the *modal depth* of a formula φ is the nesting number of modal operators appearing in φ .

We continue by defining some relevant types of formulas. Let an epistemic state $\langle \mathcal{M}, s \rangle$ be given, and a logical language LAN , where LAN is either L or L^C .

The *factual description* $\sigma_s \in \mathsf{LAN}$ of state s in \mathcal{M} is the conjunction of the values of all variables in s : Let $\sigma_p = p$ if $s \in \llbracket p \rrbracket_{\mathcal{M}}$ and else $\sigma_p = \neg p$, then $\sigma_s := \bigwedge \sigma_p$.

A *characteristic formula* $\chi_{\langle \mathcal{M}, s \rangle}$, or, alternatively, a *description of the epistemic state* $\langle \mathcal{M}, s \rangle$, is a LAN formula that is true in $\langle \mathcal{M}, s \rangle$ and such that $\langle \mathcal{N}, t \rangle \models \chi_{\langle \mathcal{M}, s \rangle}$ implies $\langle \mathcal{N}, t \rangle \rightleftharpoons \langle \mathcal{M}, s \rangle$. Similarly a *description of an epistemic model* \mathcal{M} is a LAN formula $\chi_{\mathcal{M}}$ true in the model \mathcal{M} and such that for all \mathcal{N} , $\mathcal{N} \models \chi_{\mathcal{M}}$ implies $\mathcal{M} \rightleftharpoons \mathcal{N}$.

A *distinguishing formula* between two subsets S, S' of an epistemic model \mathcal{M} is a LAN formula $\delta_{S, S'}$ such that $S \subseteq \llbracket \delta_{S, S'} \rrbracket_{\mathcal{M}}$ whereas $S' \cap \llbracket \delta_{S, S'} \rrbracket_{\mathcal{M}} = \emptyset$. If $S = \{x\}$ and $S' = \{y\}$ we write $\delta_{x, y}$ and we say that $\delta_{x, y}$ distinguishes x from y in \mathcal{M} ; and for a formula that distinguishes a state x from all other (non-bisimilar) states in the model \mathcal{M} we write δ_x . For a distinguishing δ_x we have that for all t in \mathcal{M} , $\langle \mathcal{M}, t \rangle \models \delta_x$ implies $\langle \mathcal{M}, t \rangle \rightleftharpoons \langle \mathcal{M}, x \rangle$.

4 Undefinability

Some (meta-) property Φ is definable over logical language LAN (where LAN is L or L^C) iff there is some $\varphi \in \mathsf{LAN}$ such that for all epistemic states $\langle \mathcal{M}, s \rangle$ we have $\langle \mathcal{M}, s \rangle \models \varphi$ iff $\langle \mathcal{M}, s \rangle$ satisfies Φ . In particular, to say that ‘being bisimilar to a finite epistemic state’ is definable over LAN means that for every $\langle \mathcal{M}, x \rangle$ with $|\mathcal{M}|$ being finite, there is some formula $\varphi_{\langle \mathcal{M}, x \rangle} \in \mathsf{LAN}$ such that, for every epistemic state $\langle \mathcal{N}, y \rangle$,

$$\langle \mathcal{N}, y \rangle \models \varphi_{\langle \mathcal{M}, x \rangle} \Leftrightarrow \langle \mathcal{M}, x \rangle \rightleftharpoons \langle \mathcal{N}, y \rangle. \quad (1)$$

Replacing \rightleftharpoons by \leq in (1) we have definability of ‘being simulated by a finite epistemic state’.

Theorem 1. *The property of being bisimilar to a finite epistemic state is not definable over L .*

Proof. See also Figure 3. Given $N < \omega$, consider the model \mathcal{S}_N consisting of $2N + 1$ states

$$|\mathcal{S}_N| = \{s_n : -N \leq n \leq N\}$$

such that $s_n \stackrel{a}{\sim} s_{n+1}$ if n is even, $s_n \stackrel{b}{\sim} s_{n+1}$ if n is odd and

$$\llbracket p \rrbracket_{\mathcal{S}_N} = \{s_{2n} \in |\mathcal{S}_N| : n < N\}.$$

Let \mathcal{E} be a model with two worlds e_0, e_1 , both indistinguishable to both agents and with $\llbracket p \rrbracket_{\mathcal{E}} = \{e_0\}$.

Then, clearly, for all N

$$\langle \mathcal{S}_N, s_0 \rangle \not\equiv \langle \mathcal{E}, e_0 \rangle \quad (2)$$

yet an easy induction shows that

$$\langle \mathcal{S}_N, s_0 \rangle \equiv_N \langle \mathcal{E}, e_0 \rangle \quad (3)$$

Now suppose there would be a formula $\varphi_{\langle \mathcal{E}, e_0 \rangle}$ such that for all states $\langle \mathcal{M}, x \rangle$, we would have $\langle \mathcal{M}, x \rangle \models \varphi_{\langle \mathcal{E}, e_0 \rangle}$ iff $\langle \mathcal{M}, x \rangle \equiv \langle \mathcal{E}, e_0 \rangle$. Let k be the modal depth of $\varphi_{\langle \mathcal{E}, e_0 \rangle}$. Since every model is bisimilar to itself, we would have $\langle \mathcal{E}, e_0 \rangle \models \varphi_{\langle \mathcal{E}, e_0 \rangle}$. Also, by (3) we would conclude $\langle \mathcal{S}_k, s_0 \rangle \models \varphi_{\langle \mathcal{E}, e_0 \rangle}$, which would then yield, by our assumption, that $\langle \mathcal{S}_k, s_0 \rangle \equiv \langle \mathcal{E}, e_0 \rangle$, which contradicts (2).

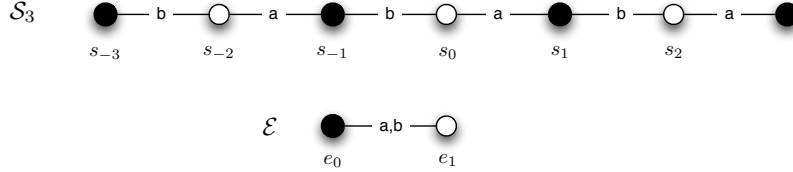


Fig. 3. The models \mathcal{E} and \mathcal{S}_3 . p -states are black.

Theorem 2. *The property of being simulated by a finite epistemic state is not definable over \mathbf{L}^C .*

Proof. Let \mathcal{E} be as in the proof of Theorem 1. We will define two epistemic states $\langle \mathcal{H}, 0 \rangle$ and $\langle \mathcal{H}', 0 \rangle$ satisfying the following properties:

1. for all $\varphi \in \mathbf{L}^C$: $\langle \mathcal{H}, 0 \rangle \models \varphi$ iff $\langle \mathcal{H}', 0 \rangle \models \varphi$.
2. $\langle \mathcal{E}, e_0 \rangle \sqsubseteq \langle \mathcal{H}', 0 \rangle$
3. not $\langle \mathcal{E}, e_0 \rangle \sqsubseteq \langle \mathcal{H}, 0 \rangle$.

Consider the two variants of a ‘spider’ (See also Figure 4.) The first of those is \mathcal{H} given by

$$|\mathcal{H}| = \{0\} \cup \{\langle n, m \rangle : 0 < |n| < m\}$$

with $\langle n, m \rangle \stackrel{a}{\sim} \langle n+1, m \rangle$ if n is odd, $\langle n, m \rangle \stackrel{b}{\sim} \langle n+1, m \rangle$ if n is even, $0 \stackrel{a}{\sim} \langle 1, m \rangle$ and $0 \stackrel{b}{\sim} \langle -1, m \rangle$ for all m and

$$\llbracket p \rrbracket_{\mathcal{H}} = \{\langle n, m \rangle : n \text{ is odd}\}.$$

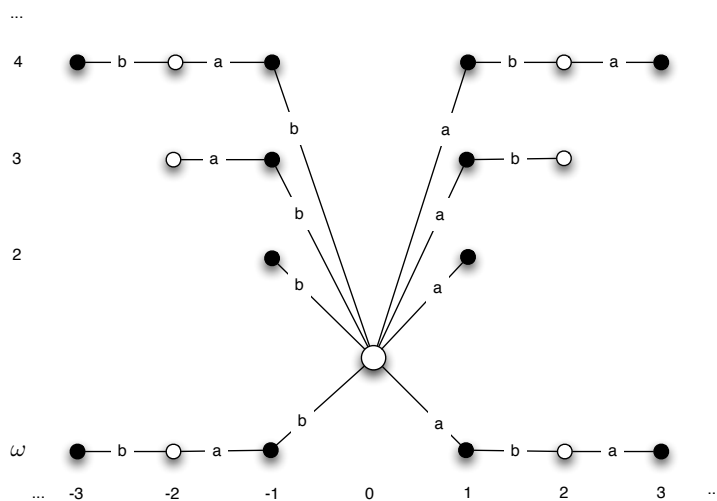


Fig. 4. Part of the spider \mathcal{H}' ($0 < |n| < 5$ and $m \in \{2, 3, 4, \omega\}$). p -states are black, $\neg p$ -states are white. The large state at the center is 0. The two bottom legs are infinite; \mathcal{H} lacks them.

Then consider the spider \mathcal{H}' , defined as \mathcal{H} but with $\{\langle n, \omega \rangle : n \in \mathbb{Z} \setminus \{0\}\}$ added (this set represents two ‘infinite legs’ of the spider).

Then, $\langle \mathcal{H}, 0 \rangle$ and $\langle \mathcal{H}', 0 \rangle$ satisfy the same set of L^C formulas; we omit the proof, which uses common knowledge games (cf. [11]).

Now, being simulated by $\langle \mathcal{E}, e_0 \rangle$ is equivalent to there being an infinite sequence $\langle s_n : n \in \mathbb{Z} \rangle$ of possibly repeating states such that $s_0 = 0$, $s_{n+1} \stackrel{a}{\sim} s_n$ if n is even, $s_{n+1} \stackrel{b}{\sim} s_n$ if n is odd, and s_n satisfies p if and only if n is odd. To see this, suppose that \mathcal{N} is some model, and $S \subseteq |\mathcal{E}| \times |\mathcal{N}|$ is a non-empty simulation. Then, there is some $s_0 \in |\mathcal{N}|$ such that $e_0 S s_0$. Now, because S is a simulation and $e_1 \stackrel{a}{\sim}_{\mathcal{E}} e_0$, there is $s_1 \stackrel{a}{\sim} s_0$ such that $e_1 S s_1$; similarly, $e_1 \stackrel{b}{\sim}_{\mathcal{E}} e_0$, so there is $s_{-1} \stackrel{b}{\sim} s_0$ such that $e_1 S s_{-1}$. Now, to find s_2 we note that $e_1 \stackrel{b}{\sim}_{\mathcal{E}} e_0$, so because S is a simulation there must be $s_2 \stackrel{b}{\sim} s_1$ with $e_0 S s_2$.

Continuing in this fashion (in both directions) we obtain s_n for each n . Note that the actual point may be repeated, but this is not important.

Now, some inspection will show that \mathcal{H}' provides such a sequence (on its ω -branch). More precisely, we define a simulation $S \subseteq |\mathcal{E}| \times |\mathcal{H}'|$ with $e_0 S 0$ and $e_i S \langle n, \omega \rangle$ if and only if $i \equiv n \pmod{2}$.

However, \mathcal{H} does not have such an infinite sequence, so $\langle \mathcal{E}, e_0 \rangle \preceq \langle \mathcal{H}', 0 \rangle$ but $\langle \mathcal{E}, e_0 \rangle \not\preceq \langle \mathcal{H}, 0 \rangle$, as claimed.

5 Definability

When describing a state s in a modal structure using a formula χ_s , what matters is the factual description σ_s of s , and a description χ_t of every accessible state t together with a statement that nothing else is accessible. So the basic pattern for state s is $\sigma_s \wedge \bigwedge_{Rst} M\chi_t \wedge K \bigvee_{Rst} \chi_t$, where, in turn, each χ_t has a pattern similar to χ_s . This is the basis of the 1970s Jankov-Fine construction to describe finite modal tree structures [3]. The unwinding of an $S5$ -structure is an infinite tree, where this does not work, unless we have infinitary modal operators or an infinitary language at our disposition. In [2] the authors show that any model $\langle M, s \rangle$ has a characteristic formula in the modal language with infinitary conjunctions and disjunctions. In the subsequent [9] an alternative route is taken via infinitary modal operators. In an older setting, [4] tackle the issue for CTL and CTL* and prove characterization of finite models in CTL. The modus operandi in [2, 9] is to introduce fresh variables p_s for a given modal structure, one for each state s , and describe the fixpoint using these fresh variables p_s . This serves to characterize modal structures but at the price of going to an extended language with an infinite stock of variables. The *explicit* purpose of these fresh variables is to make each state unique. The *implicit* justification for this procedure is that it does not matter if we change a model structure in the value of ‘irrelevant’ variables (in an $S5$ setting one could imagine a variable to be irrelevant if it is commonly known to be true or false, so that—psychologizing—it can be removed from the vocabulary in which agents reason about their uncertainty). But this approach does not serve our present goals, for two reasons: firstly, we want to give a characteristic formula in a given logical language, i.e., with a given set of propositional variables, and secondly, the procedure of introducing fresh variables makes bisimilar states s and t non-bisimilar, even at the **Atoms** level.

So, to distinguish a state from other states, state descriptions σ_s are too weak but fresh variables p_s are too strong. There is a way in between. Given a finite multi-agent $S5$ model, iterating the (multi-agent version of the) Jankov-Fine construction above up to the number of states in a model, for each s we can construct a distinguishing formula δ_s in \mathcal{L} . Clearly, this construction may be costly. Van Benthem also mentions the distinguishing formulae δ_s in [9], in the setting of PDL. Another result in [9] is that every finite $\langle M, s \rangle$ has a characteristic formula in PDL with the Kleene *, using the existence of such distinguishing formulas δ_s . In [9] in fact only the unimodal case is considered, over arbitrary K-models. This can be generalized to a set of PDL-action labels a, b, \dots ; we fill in the details here (see also [10, 12]). Given a finite set of action labels $a_1, \dots, a_n = A$, we can see the corresponding dynamic modal operators $[a_1], \dots, [a_n]$ also as epistemic modal operators K_{a_1}, \dots, K_{a_n} ; and instead of the Kleene-* applied to the *choice* between *all* these operators (a crucial detail!), i.e., $(a_1 + \dots + a_n)^*$, interpreted by the accessibility relation $(R_{a_1} \cup \dots \cup R_{a_n})^*$, we take the common knowledge operator C_A . Given that translation, we can adapt [9, Proposition 3].

Lemma 1. *Each finite epistemic state $\langle \mathcal{M}, s \rangle$ is distinguished from all other (non-bisimilar) points in \mathcal{M} by a formula δ_s in epistemic logic **without** common knowledge.*

Proof. For any finite model \mathcal{M} , the following informal algorithm partitions its domain $S_{\mathcal{M}}$ into its bisimulation equivalence classes, with a distinguishing formula for each class. We give the construction for a single agent and for a finite set of variables P , the multi-agent case is similar. The algorithm consists of a finite iteration. For step 0, consider the state descriptions σ_s for all states $s \in S_{\mathcal{M}}$, and refine the domain into subsets with the same state description. Before we show step $n + 1$ given step n , let us show step 1. For step 1, we proceed with the non-singleton subsets. Consider some such subset $S' \subseteq S_{\mathcal{M}}$. For each $s \in S'$, consider the formulas $\sigma_s \wedge \bigwedge_{t \sim s} M\sigma_t \wedge K \bigvee_{t \sim s} \sigma_t$, and refine S' into subsets satisfying the same such formula. Formally, for step 0, let $\delta_s^0 := \sigma_s$; and for step $n + 1$, we get $\delta_s^{n+1} := \sigma_s \wedge \bigwedge_{t \sim s} M\delta_t^n \wedge K \bigvee_{t \sim s} \delta_t^n$. Proceed iteratively until all sets are singletons or until a maximum of $|S_{\mathcal{M}}|$ steps is reached. As the outcome of this process, each state is accompanied by a distinguishing formula δ_s , where $\delta_s = \delta_s^m$ for some $m \leq |\mathcal{M}|$. Note that all δ_s are in \mathcal{L} . In this construction, bisimilar states will end up in the same subset.¹

Theorem 3. *Each finite epistemic model \mathcal{M} is characterised by a formula in epistemic logic **without** common knowledge; that is, given finite \mathcal{M} there exists a formula $\chi_{\mathcal{M}} \in \mathcal{L}$ such that, for every (possibly infinite) epistemic model \mathcal{N} , $\mathcal{N} \models \chi_{\mathcal{M}}$ if and only if $\mathcal{M} \Leftrightarrow \mathcal{N}$.*

Proof. Let \mathcal{M} be finite, and without loss of generality, suppose it contraction-minimal, i.e., \mathcal{M} is not bisimilar to a smaller model. Also assume that it is generated. Define

$$\begin{aligned} \text{Forth}_x &= \bigwedge_{a \in A} \bigwedge_{y \overset{a}{\sim} x} M_a \delta_y \text{ and } \text{Back}_x = \bigwedge_{a \in A} K_a \bigvee_{y \overset{a}{\sim} x} \delta_y; \\ \chi_{\mathcal{M}} &= \bigvee_{x \in |\mathcal{M}|} \delta_x \wedge \bigwedge_{x \in |\mathcal{M}|} (\delta_x \rightarrow \text{Forth}_x \wedge \text{Back}_x) \end{aligned} \quad (4)$$

Now the following are equivalent, for any $\langle \mathcal{N}, t \rangle$:

- $\mathcal{M} \Leftrightarrow \mathcal{N}$
- $\mathcal{N} \models \chi_{\mathcal{M}}$.

Since $\mathcal{M}, s \models \chi_{\mathcal{M}}$, any bisimilar model \mathcal{N} also satisfies $\chi_{\mathcal{M}}$. For the converse, If $\mathcal{N} \models \chi_{\mathcal{M}}$, then we can define a relation

$$B \subseteq |\mathcal{M}| \times |\mathcal{N}|$$

¹ The maximum $|S_{\mathcal{M}}|$ of the iteration is also mentioned as Facts 7 and 8 in [9, p.30]; but no actual construction is given there. A similar maximum can be construed in the CTL setting of the older publication [4].

given by $x B y$ if and only if $y \in \llbracket \delta_x \rrbracket_{\mathcal{N}}$. Then, it remains to be checked that B is a bisimulation; indeed, if $x B y$ and $x' \overset{a}{\sim} x$, then we have that y satisfies δ_x and, since

$$\langle \mathcal{N}, t \rangle \models \delta_x \rightarrow \text{Forth}_x$$

(use $\chi_{\mathcal{M}}$), it follows that y satisfies $M_a \delta_{x'}$, and thus there is $y' \overset{a}{\sim} y$ such that y' satisfies $\delta_{x'}$, i.e., $x' B y'$. Thus B satisfies **Forth**.

A similar argument shows that B satisfies **Back** as well, and hence it is a bisimulation.

Compare this to the following, proven in [10]:

Theorem 4. *Given a finite model \mathcal{M} and $x \in |\mathcal{M}|$, there is a formula $\chi_{\langle \mathcal{M}, x \rangle} \in \mathbb{L}^C$ such that, given any finite state $\langle \mathcal{N}, y \rangle$, $\langle \mathcal{N}, y \rangle \models \chi_{\langle \mathcal{M}, x \rangle}$ if and only if $\langle \mathcal{N}, y \rangle \Leftrightarrow \langle \mathcal{M}, x \rangle$.*

Note that, by Theorem 1, the use of common knowledge is essential. Indeed, common knowledge is used as follows:

$$\chi_{\langle \mathcal{M}, s \rangle} = \delta_s \wedge C_A \bigwedge_{x \in |\mathcal{M}|} (\delta_x \rightarrow \text{Forth}_x \wedge \text{Back}_x) \quad (5)$$

We observe that common knowledge is used rather weakly, namely in the form of a single occurrence of that operator. Compare this with Theorem 3 above, that does not have the common knowledge operator. We further point out the resemblance with the distinguishing formula construction in the proof of Lemma 1. There, we use the Yankov-Fine construction iteratively up to the size of the finite model, and with building units the state descriptions. Whereas for the characteristic formula we use the Yankov-Fine construction as a conjunction over all states in the model, closed by common knowledge, and with building units the distinguishing formulas.

Now we turn our attention to a setting in which simulability does become definable:

Definition 3 (Well-multifounded models). *An epistemic model \mathcal{M} is well-multifounded if there does not exist an infinite sequence of states $s_n \overset{a(n)}{\sim} s_{n+1}$ such that for all n , $s_n \neq s_{n+1}$ and $a(n) \neq a(n+1)$.*

The model \mathcal{M}_1 from Figure 1 is well-multifounded, the models \mathcal{M}_2 and \mathcal{M}_3 from Figure 2 are not.

Theorem 5. *If \mathcal{M} is well-multifounded and finite and $s \in |\mathcal{M}|$, then simulability by $\langle \mathcal{M}, s \rangle$ is definable over \mathbb{L} .*

Proof. Suppose that \mathcal{M} is well-multifounded and finite. We first note that given two distinct states $s, t \in |\mathcal{M}|$, there is at most one agent $a \in A$ such that $s \overset{a}{\sim} t$; otherwise, this would immediately give us an infinite loop s, t, s, t, s, \dots violating well-multifoundedness.

Now, pick $s_* \in |\mathcal{M}|$. We will show, by induction on the size of \mathcal{M} , that there is a formula $\text{Sim}(\langle \mathcal{M}, s_* \rangle)$ defining the property of being simulated by $\langle \mathcal{M}, s_* \rangle$.

The base case, when \mathcal{M} has no states, is vacuously true.

Suppose, then, that a formula $\text{Sim}(\langle \mathcal{Y}, s \rangle)$ exists defining the property of being simulated by $\langle \mathcal{Y}, t \rangle$ whenever \mathcal{Y} is well-multifounded and has strictly less states than \mathcal{M} .

Consider

$$\mathcal{N} = \mathcal{M} \upharpoonright (|\mathcal{M}| \setminus \{s_*\}),$$

i.e., the model obtained by deleting s_* from \mathcal{M} .

Because \mathcal{N} is smaller than \mathcal{M} , for every $t \in |\mathcal{N}|$ there is a formula $\text{Sim}(\langle \mathcal{N}, t \rangle)$ which defines simulability by $\langle \mathcal{N}, t \rangle$; that is, such that given any epistemic model \mathcal{X} and $x \in |\mathcal{X}|$, $\langle \mathcal{N}, t \rangle \sqsubseteq \langle \mathcal{X}, x \rangle$ if and only if $\mathcal{X}, x \models \text{Sim}(\langle \mathcal{N}, t \rangle)$.

Recall that $\sigma(s)$ is the conjunction of all literals (propositional variables or their negation) which are true on s . Define

$$\text{Sim}(\langle \mathcal{M}, s_* \rangle) = \sigma(s_*) \wedge \bigwedge_{a \in A} \bigwedge_{t \stackrel{a}{\sim}_{\mathcal{M}} s_*} \text{Sim}(\langle \mathcal{N}, t \rangle).$$

We claim that $\text{Sim}(\langle \mathcal{M}, s_* \rangle)$ defines being simulated by $\langle \mathcal{M}, s_* \rangle$ over the class of epistemic models, that is, given any state $\langle \mathcal{X}, x \rangle$ we have that

$$\langle \mathcal{M}, s_* \rangle \sqsubseteq \langle \mathcal{X}, x \rangle \Leftrightarrow \langle \mathcal{X}, x \rangle \models \text{Sim}(\langle \mathcal{M}, s_* \rangle).$$

We consider each direction separately.

Left-to-right. Assume that $\langle \mathcal{M}, s_* \rangle \sqsubseteq \langle \mathcal{X}, x \rangle$, so that there is a simulation

$$S \subseteq |\mathcal{M}| \times |\mathcal{X}|$$

with $s_* S x$. We need to show that $\mathcal{X}, x \models \text{Sim}(\langle \mathcal{M}, s_* \rangle)$.

Clearly $x \in \llbracket \sigma(s_*) \rrbracket_{\mathcal{X}}$. Now, if $t \stackrel{a}{\sim}_{\mathcal{M}} s_*$, because S is a simulation there is $y \stackrel{a}{\sim}_{\mathcal{X}} x$ such that $t S y$.

We need to show that $y \in \llbracket \text{Sim}(\langle \mathcal{N}, t \rangle) \rrbracket_{\mathcal{X}}$ to establish our claim; but by our induction hypothesis, it suffices to observe that $S \upharpoonright |\mathcal{N}|$ is a simulation (the ‘forth’ condition is clearly preserved by submodels), and by assumption the existence of such a simulation implies that y satisfies $\text{Sim}(\langle \mathcal{N}, t \rangle)$.

Looking at the definition of $\text{Sim}(\langle \mathcal{M}, s_* \rangle)$, it follows that $x \in \llbracket \text{Sim}(\langle \mathcal{M}, s_* \rangle) \rrbracket_{\mathcal{X}}$, as desired.

Right-to-left. Suppose that $x_* \in \llbracket \text{Sim}(\langle \mathcal{M}, s_* \rangle) \rrbracket_{\mathcal{X}}$; we need to construct a simulation $S \subseteq |\mathcal{M}| \times |\mathcal{X}|$ such that $s_* S x_*$. In fact, for our inductive argument to work it is convenient to define S so that it is a function; let us call a simulation which is a function an *embedding*. Let B be the set of all agents a such that there is $t \neq s_*$ with $t \stackrel{a}{\sim}_{\mathcal{M}} s_*$. For each $a \in B$ pick a fixed representative $d(a)$ with the property that $d(a) \stackrel{a}{\sim}_{\mathcal{M}} s_*$.

Let \mathcal{N}_a define the submodel of \mathcal{N} generated by $d(a)$. We claim:

1. each \mathcal{N}_a is well-multifounded.
(Since well-multifoundedness is preserved under submodels);
2. \mathcal{N}_a does not depend on the specific choice of $d(a)$; that is, for all $t \stackrel{a}{\sim}_{\mathcal{M}} s_*$, the model \mathcal{N}_a equals the submodel of \mathcal{N} generated by t .
(If $t \stackrel{a}{\sim}_{\mathcal{M}} s_*$, by transitivity we have that $t \stackrel{a}{\sim} d(a)$, so $t \in |\mathcal{N}_a|$; similarly, $d(a)$ is contained in the submodel generated by t , so the two are equal.)
3. $|\mathcal{N}_a| \cap |\mathcal{N}_b| = \emptyset$ whenever $a \neq b$.
(If $t \in |\mathcal{N}_a| \cap |\mathcal{N}_b|$, we have a path connecting $d(a)$ to t and another connecting $d(b)$ to t (this is the definition of lying in the generated submodel). ‘Pruning’ the path if necessary, we can assume that points do not repeat and the agents connecting consecutive points are distinct.
These two paths give us a ‘loop’ which begins on s_* and passes through $d(a)$, then t , then $d(b)$ and back to s_* ; we can then run this loop infinitely many times to obtain a sequence violating well-multifoundedness. We conclude that there is no such t .)

For every $a \in B$ and $t \stackrel{a}{\sim} s_*$, we have $x_* \in \llbracket M_a \text{Sim}(\langle \mathcal{N}, t \rangle) \rrbracket_{\mathcal{X}}$ (by $\text{Sim}(\langle \mathcal{M}, s_* \rangle)$); in particular,

$$x_* \in \llbracket M_a \text{Sim}(\langle \mathcal{N}, d(a) \rangle) \rrbracket_{\mathcal{X}}.$$

By induction we can assume that there is $y(a) \stackrel{a}{\sim}_{\mathcal{X}} x_*$ and an embedding $S_a \subseteq |\mathcal{N}_a| \times |\mathcal{X}|$ such that $d(a) S_a(d(a)) = y(a)^2$. Define

$$S = \{\langle s_*, x_* \rangle\} \cup \bigcup_{a \in B} S_a.$$

We claim that S is an embedding; clearly it preserves propositional variables; it remains to check that the ‘forth’ condition holds. It is easy to see that S is indeed a function, since we have only added the pair $\langle s_*, x_* \rangle$ to a disjoint union of functions (with disjoint domains). Thus the ‘forth’ condition becomes: if $s \stackrel{a}{\sim}_{\mathcal{M}} t$ then $S(s) \stackrel{a}{\sim}_{\mathcal{X}} S(t)$.

To prove this, pick $s \stackrel{a}{\sim}_{\mathcal{M}} t$ and suppose that $x = S(s)$. There are essentially three cases we must consider. (1) $s, t \in \mathcal{N}$. They both lie in some \mathcal{N}_a , since these models are generated; but then S_a is an embedding, so $S_a(s) \stackrel{a}{\sim}_{\mathcal{X}} S_a(t)$. (2) $s = s_*$, so that $S(s) = x_*$. By transitivity of $\stackrel{a}{\sim}_{\mathcal{M}}$ we have that $t \stackrel{a}{\sim}_{\mathcal{M}} d(a)$, hence $S_a(t) \stackrel{a}{\sim}_{\mathcal{X}} y(a)$ and, by transitivity of $\stackrel{a}{\sim}_{\mathcal{X}}$, we have that $S_a(t) \stackrel{a}{\sim}_{\mathcal{X}} x_*$, which is what we needed. (3) $t = s_*$. Then $s \in |\mathcal{N}_a|$, and because these sets are disjoint it is the only agent for which this holds. Thus $S(s) = S_a(s)$. Now, $s \stackrel{a}{\sim}_{\mathcal{M}} s_* \stackrel{a}{\sim}_{\mathcal{M}} d(a)$, so by transitivity $s \stackrel{a}{\sim}_{\mathcal{M}} d(a)$ and, because S_a is a simulation, $S_a(s) \stackrel{a}{\sim}_{\mathcal{X}} y(a) \stackrel{a}{\sim}_{\mathcal{X}} x_*$, as desired.

6 Conclusion and further research

When building multi-agent systems applications, in case you want to execute some dynamics, you often want to apply the action ‘right here in the model’.

² Properly speaking, we begin with an embedding S'_a between \mathcal{N} and \mathcal{X} and restrict it to \mathcal{N}_a .

Unless you have nominals, you can enforce this by preconditions that distinguish that state from other states. This can be done in logic without common knowledge. However, there are different cases wherein you want to pin down the current state including all its epistemic aspects. ‘Only under *exactly* these conditions, this action will apply’. Then you need a characteristic formula and *that* can only be done in epistemic logic with common knowledge. This significant distinction may not always be observed. We hope that our results, already known in the communities of PDL and CTL, may therefore contribute to logical hygiene in epistemic applications.

Computing characteristic formulas employing distinguishing formulas may be costly. For specific model classes the characteristic formulas may be simpler. Static interpreted systems [5] consist of a domain of global states, where each global state is a collection of local states, one for each agent. If agents only know their local state, the description of the global state (i.e., the description of the valuation) is a distinguishing formula in the Kripke model for that interpreted system. (See also [7] on the relation between interpreted systems wherein agents *only* know their local state, and Kripke models, by way of ‘hypercubes’.) Also, for such systems, the characteristic formula has a special shape wherein factual (purely propositional), positive, and negative knowledge can be separated as different parts of the formula. The factual part says that one of all global states in the system must be the case (this is therefore a large disjunction of state descriptions); the positive part says that each agent knows her local state, and the negative part sums up the ignorance. This approach is followed in [12] for interpreted systems modelling card deals—but it can clearly be generalized. The constructions in the previous section, instead, consist of many conjuncts in each of which all these three types of formula are mixed. The modular form of characteristic formulas for such interpreted systems is an advantage, because when there are dynamic developments in the system it allows us to focus on the changing parts of the formula: factual and positive knowledge are always preserved, and only the negative knowledge, the ignorance, is reduced.

Here our results on simulability are also useful, since simulation is intimately related with epistemic actions. Roughly speaking, epistemic actions are events by which agents learn new information, and given an epistemic model \mathcal{M} , any submodel \mathcal{N} of \mathcal{M} (and even more: if \mathcal{N} results from executing whatever action model in \mathcal{M} , then \mathcal{N} is simulated by \mathcal{M}) models a situation where agents have more positive knowledge than they had on \mathcal{M} , since each new world on \mathcal{M} represents a new source of uncertainty for the agents. *Simulation* is the bisimulation-invariant analogue of being a submodel; that is, $\mathcal{N} \preceq \mathcal{M}$ is only slightly more general than to say that \mathcal{N} is bisimilar to a submodel of \mathcal{M} .

Thus defining submodels up to simulability helps us interiorize dynamics into the syntax by fully describing the effect of certain epistemic actions on a model. This could be used to turn a static description of a model into a dynamic one, by describing which epistemic states may result after executing those actions which are available to the agents.

For future research, it is also worthwhile to determine (and lower) the complexity of the computation of distinguishing formulas, the building stones of the characteristic formulas. Complexity worries were already uttered in [4]—but we do not know of any subsequent resolution. In the case of static interpreted systems where state descriptions are already distinguishing, this is clearly more efficient.

Acknowledgment

We thank the CLIMA reviewers for comments. Hans van Ditmarsch is also affiliated to IMSC (Institute of Mathematical Sciences Chennai), India, as associated researcher.

References

1. P. Aczel. *Non-Well-Founded Sets*. CSLI Publications, Stanford, CA, 1988. CSLI Lecture Notes 14.
2. J. Barwise and L.S. Moss. *Vicious Circles*. CSLI Publications, Stanford, 1996.
3. P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001. Cambridge Tracts in Theoretical Computer Science 53.
4. M. Browne, E. Clarke, and O. Grumberg. Characterizing Kripke structures in temporal logic. In H. Ehrig, R. Kowalski, G. Levi, and U. Montanari, editors, *TAPSOFT '87*, pages 256–270. Springer, 1987. LNCS 249.
5. R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. MIT Press, 1995.
6. D. Fernández-Duque. On the modal definability of simulability by finite transitive models. *Studia Logica*, 2011. Forthcoming.
7. A.R. Lomuscio and M.D. Ryan. On the relation between interpreted systems and Kripke models. In M. Pagnucco, W.R. Wobcke, and C. Zhang, editors, *Proceedings of the AI97 workshop on the theoretical and practical foundations of intelligent agents and agent-oriented systems*, pages 46–59, Berlin, 1998. Springer. Lecture Notes in Artificial Intelligence 1441.
8. J.-J.Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge Tracts in Theoretical Computer Science 41. Cambridge University Press, Cambridge, 1995.
9. J. van Benthem. Dynamic odds and ends. *ILLC Technical Report ML*, 1998.
10. J. van Benthem. ‘One is a lonely number’: on the logic of communication. In Z. Chatzidakis, P. Koepke, and W. Pohlers, editors, *Logic Colloquium '02*. Association for Symbolic Logic, 2006. Volume 27 of *Lecture Notes in Logic*.
11. H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer, Berlin, 2007.
12. H. van Ditmarsch, W. van der Hoek, and B.P. Kooi. Descriptions of game states. In G. Mints and R. Muskens, editors, *Logic, Games, and Constructive Sets*, pages 43–58, Stanford, 2003. CSLI Publications. CSLI Lecture Notes No. 161.