A Modal Characterization Theorem for a Probabilistic Fuzzy Description Logic

Paul Wild[1] , **Lutz Schröder**[1] , **Dirk Pattinson**[2] , **Barbara König**[3]
[1]Friedrich-Alexander-Universität Erlangen-Nürnberg
[2]Australian National University, Canberra
[3]Universität Duisburg-Essen

Abstract

The fuzzy modality *probably* is interpreted over probabilistic type spaces by taking expected truth values. The arising probabilistic fuzzy description logic is invariant under probabilistic bisimilarity; more informatively, it is non-expansive wrt. a suitable notion of behavioural distance. In the present paper, we provide a characterization of the expressive power of this logic based on this observation: We prove a probabilistic analogue of the classical van Benthem theorem, which states that modal logic is precisely the bisimulation-invariant fragment of first-order logic. Specifically, we show that every formula in probabilistic fuzzy first-order logic that is non-expansive wrt. behavioural distance can be approximated by concepts of bounded rank in probabilistic fuzzy description logic.

1 Introduction

In the representation of uncertain knowledge, one will often wish to avoid mention of exact numerical probabilities, e.g. when these are not precisely known or not relevant to the representation task at hand – as a typical example, a medical practitioner will rarely name a numerical threshold for the likelihood of a diagnosis, and instead qualify the diagnosis as, say, 'suspected' or 'probable'. This has led to efforts aimed at formalizing a modality *probably*, used alternatively to modalities 'with probability at least p' [Larsen and Skou, 1991; Heifetz and Mongin, 2001; Lutz and Schröder, 2010]. Such a formalization may be approached in a two-valued setting via qualitative axiomatizations of likelihood [Burgess, 1969; Halpern and Rabin, 1987] or via threshold probabilities [Hamblin, 1959; Herzig, 2003]. In a fuzzy setting, 'probably' leads a natural life as a fuzzy modality P, whose truth value just increases as its argument becomes more probable (this modality thus connects the otherwise well-distinguished worlds of fuzziness and probability [Lukasiewicz and Straccia, 2008]). The semantics of this operator, first defined by Zadeh [1968], interprets $\mathsf{P}\,\phi$ as the expected truth value of ϕ. It appears in various fuzzy propositional [Hájek, 2007; Flaminio and Godo, 2007], modal [Desharnais *et al.*, 1999; van Breugel and Worrell, 2005], fixpoint [Kozen, 1985; Huth and Kwiatkowska, 1997], and description logics [Schröder and Pattinson, 2011].

In the present paper, we pin down the exact expressiveness of the basic description logic of *probably*, which we briefly refer to as *probabilistic fuzzy* $\mathcal{ALC}$ or $\mathcal{ALC}(\mathsf{P})$, within a natural ambient probabilistic fuzzy first-order logic $\mathsf{FO}(\mathsf{P})$, by providing a *modal characterization theorem*. The prototype of such characterization theorems is *van Benthem's theorem* [1976], which states that (classical) modal logic is precisely the bisimulation-invariant fragment of first-order logic. It has been noted that in systems with numerical values, *behavioural pseudometrics* offer a more fine-grained measure of equivalence than two-valued bisimilarity [Giacalone *et al.*, 1990; Desharnais *et al.*, 1999; van Breugel and Worrell, 2005; Desharnais *et al.*, 2008; Baldan *et al.*, 2014]. When propositional connectives are equipped with Zadeh semantics, $\mathcal{ALC}(\mathsf{P})$ is *non-expansive* wrt. behavioural distance; we continue to refer to this property as *bisimulation invariance*. In previous work [Wild *et al.*, 2018] we have shown that *relational* fuzzy modal logic is the bisimulation-invariant fragment of fuzzy FOL, more precisely that every bisimulation-invariant fuzzy FO formula can be approximated by fuzzy modal formulae *of bounded rank*. The bound on the rank is key; without it, the statement turns into a form of the (much simpler) Hennessy-Milner theorem [Hennessy and Milner, 1985] (which classically states that non-bisimilar states in finitely branching systems can be distinguished by modal formulae), and indeed does not need to assume FO definability of the given bisimulation-invariant property [van Breugel and Worrell, 2005]. Here, we establish a corresponding result for the rather more involved probabilistic setting: We show that *every bisimulation-invariant formula in probabilistic fuzzy FOL can be approximated in bounded rank in probabilistic fuzzy* $\mathcal{ALC}$. This means not only that, up to approximation, $\mathcal{ALC}(\mathsf{P})$ is as powerful as $\mathsf{FO}(\mathsf{P})$ on bisimulation-invariant properties, but also that $\mathcal{ALC}(\mathsf{P})$ provides effective syntax for bisimulation-invariant $\mathsf{FO}(\mathsf{P})$, which $\mathsf{FO}(\mathsf{P})$ itself does not [Otto, 2006].

Proofs are mostly omitted or only sketched; full proofs are in the appendix.

Related Work There is widespread interest in modal characterization theorems in modal logic [Dawar and Otto, 2005], database theory [Figueira *et al.*, 2015], concurrency [Janin and Walukiewicz, 1995; Carreiro, 2015], and AI [Sturm and

Wolter, 2001; Wild and Schröder, 2017; Wild *et al.*, 2018]. The overall structure of our proof builds partly on that of our modal characterization theorem for relational fuzzy modal logic [Wild *et al.*, 2018] (in turn based ultimately on a strategy due to Otto [2004]) but deals with a much more involved logic, which instead of just the lattice structure of the unit interval involves its full arithmetic structure, via the use of probabilities and expected values, necessitating, e.g., the use of Kantorovich-Rubinstein duality. Notable contributions of our proof include new forms of probabilistic bisimulation games up-to-ϵ (different from games introduced by Desharnais et al. [2008], which characterize a different metric) and Ehrenfeucht-Fraïssé games, related to two-valued games considered in the context of topological FOL [Makowsky and Ziegler, 1980]. (For lack of space, we omit discussion of quantitative Hennessy-Milner type results beyond the mentioned result by van Breugel and Worrell [2005].)

$\mathsf{FO}(\mathsf{P})$ may be seen as a fuzzy variant of Halpern's [1990] type-1 (i.e. statistical) two-valued probabilistic FOL, and uses a syntax related to coalgebraic predicate logic [Litak *et al.*, 2018] and, ultimately, Chang's *modal predicate logic* [Chang, 1973]. Van-Benthem style theorems for two-valued coalgebraic modal logic [Schröder *et al.*, 2017] instantiate to two-valued probabilistic modal logic, then establishing expressibility of bisimulation-invariant probabilistic FO formulae by probabilistic modal formulae with infinite conjunction but of bounded rank, in an apparent analogy to bounded-rank approximation in the fuzzy setting.

2 Fuzzy Probabilistic Logics

We proceed to introduce the logics featuring in our main result. We fix (w.l.o.g., finite) sets $\mathsf{N_C}$ of *atomic concepts* and $\mathsf{N_R}$ of *roles*; *concepts* C, D of *quantitative probabilistic* $\mathcal{ALC}$ ($\mathcal{ALC}(\mathsf{P})$) are defined by the grammar

$$C, D ::= q \mid A \mid C \ominus q \mid \neg C \mid C \sqcap D \mid \mathsf{P}\, r.\, C$$

where $q \in \mathbb{Q} \cap [0,1]$, $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. The intended reading of P is 'probably'; we give examples below. Slightly deviating from standard practice, we define the *rank* $\mathsf{rk}(C)$ of a concept C as the maximal nesting depth of the P *and atomic concepts* in C; e.g. $\mathsf{rk}((\mathsf{P}\, r.\, \mathsf{P}\, s.\, A) \sqcap (\mathsf{P}\, r.\, B)) = 3$. We denote the set of all concepts of rank at most n by $\mathcal{ALC}(\mathsf{P})_n$.

Concepts are interpreted over probabilistic structures to which we neutrally refer as *interpretations* or, briefly, *models*. We allow infinite models but restrict to discrete probability distributions over successors at each state. A model

$$\mathcal{I} = (\Delta^{\mathcal{I}}, (A^{\mathcal{I}})_{A \in \mathsf{N_C}}, (r^{\mathcal{I}})_{r \in \mathsf{N_R}})$$

consists of a *domain* $\Delta^{\mathcal{I}}$ of *states* or *individuals*, and interpretations $A^{\mathcal{I}} \colon \Delta^{\mathcal{I}} \to [0,1]$, $r^{\mathcal{I}} \colon \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \to [0,1]$ of atomic concepts A and roles r such that for each $a \in \Delta^{\mathcal{I}}$, the map

$$r_a \colon \Delta^{\mathcal{I}} \to [0,1], \quad r_a(a') = r^{\mathcal{I}}(a, a')$$

is either zero or a probability mass function on $\Delta^{\mathcal{I}}$, i.e.

$$\textstyle\sum_{a' \in \Delta^{\mathcal{I}}} r_a(a') \in \{0,1\}$$

(implying that the *support* $\{a' \in \Delta^{\mathcal{I}} \mid r_a(a') > 0\}$ of r_a is at most countable). We call a state a *r-blocking* if $\sum_{a' \in \Delta^{\mathcal{I}}} r_a(a') = 0$. At non-blocking states a, r_a thus acts as a probabilistic accessibility relation; we abuse r_a to denote also the probability measure defined by r_a.

The interpretation $C^{\mathcal{I}} \colon \Delta^{\mathcal{I}} \to [0,1]$ of concepts is defined recursively, extending that of atomic concepts, by

$$\begin{aligned} q^{\mathcal{I}}(a) &= q \\ (C \ominus q)^{\mathcal{I}}(a) &= \max(C^{\mathcal{I}}(a) - q, 0) \\ (\neg C)^{\mathcal{I}}(a) &= 1 - C^{\mathcal{I}}(a) \\ (C \sqcap D)^{\mathcal{I}}(a) &= \min(C^{\mathcal{I}}(a), D^{\mathcal{I}}(a)) \\ (\mathsf{P}\, r.\, C)^{\mathcal{I}}(a) &= \mathrm{E}_{r_a}(C^{\mathcal{I}}) = \textstyle\sum_{a' \in \Delta^{\mathcal{I}}} r_a(a') \cdot C^{\mathcal{I}}(a') \end{aligned}$$

At non-blocking a, $(\mathsf{P}\, r.\, C)^{\mathcal{I}}(a)$ is thus the expected truth value of C for a random r-successor of a. We define disjunction $\sqcup$ as the dual of $\sqcap$ as usual, so $\sqcup$ takes maxima. We use Zadeh semantics for the propositional operators, which will later ensure non-expansiveness wrt. behavioural distance; see additional comments in Section 7.

Up to minor variations, our models correspond to Markov chains or, in an epistemic reading, *type spaces* (e.g. [Heifetz and Mongin, 2001]). The logic $\mathcal{ALC}(\mathsf{P})$ was considered (with Łukasiewicz semantics) by Schröder and Pattinson [2011], and resembles van Breugel and Worrell's quantitative probabilistic modal logic [2005]. E.g., in a reading of $\Delta^{\mathcal{I}}$ as consisting of real-world individuals, the concept

$$\mathsf{Loud} \sqcap \mathsf{P}\, \mathsf{hasSource}.\, (\mathsf{Large} \sqcap \mathsf{P}\, \mathsf{hasMood}.\, \mathsf{Angry})$$

describes noises you hear in your tent at night as being loud and probably coming from the large and probably angry animal whose shadow just crossed the tent roof. (In this view, P can be usefully combined with crisp or fuzzy relational modalities, using off-the-shelf compositionality mechanisms [Schröder and Pattinson, 2011].) In an epistemic reading where the elements of $\Delta^{\mathcal{I}}$ are possible worlds, and the roles are understood as epistemic agents, the concept

$$\neg\mathsf{GoodHand} \sqcap \mathsf{P}\, \mathsf{player}.\, \mathsf{P}\, \mathsf{opponent}.\, \mathsf{GoodHand}$$

denotes the degree to which player believes she is successfully bluffing by letting opponent overestimate player's hand.

For readability, we will restrict the technical treatment to a single role r, omitted in the syntax, from now on, noting that covering multiple roles amounts to no more than additional indexing. As the first-order correspondence language of quantitative probabilistic $\mathcal{ALC}$ we introduce *quantitative probabilistic first-order logic* ($\mathsf{FO}(\mathsf{P})$), with *formulae* $\phi, \psi, \dots$ defined by the grammar

$$\begin{aligned} \phi, \psi ::= q \mid A(x) \mid x = y \mid \phi \ominus q \mid \neg\phi \mid \phi \sqcap \psi \mid \exists x.\, \phi \\ \mid x\mathsf{P}\lceil y : \phi \rceil \qquad (q \in \mathbb{Q} \cap [0,1], A \in \mathsf{N_C}) \end{aligned}$$

where x and y range over a fixed countably infinite reservoir of *variables*. The reading of $x\mathsf{P}\lceil y : \phi\rceil$ is the expected truth value of ϕ at a random successor y of x. (In particular, when ϕ is crisp, then $x\mathsf{P}\lceil y : \phi\rceil$ is just the probability of y satisfying ϕ, similar to the weights $w_y(\phi)$ in Halpern's type-1 probabilistic FOL [1990].) We have the expected notions of free and bound variables, under the additional proviso that y (but not x!) is bound in $x\mathsf{P}\lceil y : \phi\rceil$. The *(quantifier)* *rank* $\mathsf{qr}(\phi)$ of a formula ϕ is the maximal nesting depth of the variable-binding operators $\exists$ and P and propositional atoms A in ϕ; e.g. $\exists x.\, x\mathsf{P}\lceil y : A(y)\rceil$ has rank 3.

Given a model $\mathcal{I} = (\Delta^{\mathcal{I}}, (A^{\mathcal{I}})_{A\in\mathsf{N_C}}, r^{\mathcal{I}})$ and a vector $\bar{a} = (a_1,\dots,a_n) \in (\Delta^{\mathcal{I}})^n$ of values, the semantics of the logic assigns a truth value $\phi(\bar{a}) \in [0,1]$ to a formula $\phi(x_1,\dots,x_n)$ with free variables at most $x_1,\dots,x_n$. We define $\phi(\bar{a})$ recursively by essentially the same clauses as in $\mathcal{ALC}(\mathsf{P})$ for the propositional constructs, and

$$A(x_i)(\bar{a}) = A^{\mathcal{I}}(a_i)$$
$$(\exists x_0.\, \phi(x_0,x_1,\dots,x_n))(\bar{a}) = \textstyle\bigvee_{a_0\in\Delta^{\mathcal{I}}} \phi(a_0,a_1,\dots,a_n)$$
$$(x_i\mathsf{P}\lceil y : \phi(y,x_1,\dots,x_n)\rceil)(\bar{a}) = \mathrm{E}_{r_{a_i}}(\phi(\,\cdot\,,a_1,\dots,a_n))$$

where $\bigvee$ takes suprema. Moreover, equality is two-valued, i.e. $(x_i = x_j)(\bar{a})$ is 1 if $a_i = a_j$, and 0 otherwise.

E.g. the formula $x\mathsf{P}\lceil z : z = y\rceil$ ('the successor of x is probably y') denotes the access probability from x to y, $x\mathsf{P}\lceil z : z\mathsf{P}\lceil w : w = y\rceil\rceil$ the probability of reaching y from x in two independently distributed steps, and $\exists y.\, x\mathsf{P}\lceil z : z = y\rceil$ the probability of the most probable successor of x.

We have a *standard translation* ST_x from $\mathcal{ALC}(\mathsf{P})$ into $\mathsf{FO}(\mathsf{P})$, indexed over a variable x naming the current state. Following Litak et al. [2018], we define ST_x recursively by

$$\mathsf{ST}_x(A) = A(x)$$
$$\mathsf{ST}_x(\mathsf{P}C) = x\mathsf{P}\lceil y : \mathsf{ST}_y(C)\rceil,$$

and by commutation with all other constructs.

Lemma 2.1. *For every $\mathcal{ALC}(\mathsf{P})$-concept C and state a, $C(a) = \mathsf{ST}_x(C)(a)$.*

ST thus identifies $\mathcal{ALC}(\mathsf{P})$ as a fragment of $\mathsf{FO}(\mathsf{P})$.

3 Behavioural Distances and Games

We next discuss several notions of behavioural distance between states: via fixed point iteration à la Wasserstein/Kantorovich, via games and via the logic. We focus mostly on depth-n distances. Only for one version, we define also the unbounded distance, which will feature in the modal characterization result. We show in Section 4 that at finite depth, all these distances coincide. It has been shown in previous work [Desharnais *et al.*, 2004; van Breugel and Worrell, 2005] that the unbounded-depth distances defined via Kantorovich fixed point iteration and via the logic, respectively, coincide in very similar settings; such results can be seen as probabilistic variants of the Hennessy-Milner theorem.

We recall standard notions on pseudometric spaces:

Definition 3.1 (Pseudometric spaces, non-expansive maps)**.** A *(bounded) pseudometric* on a set X is a function $d\colon X \times X \to [0,1]$ such that for $x,y,z \in X$, the following axioms hold: $d(x,x) = 0$ (*reflexivity*), $d(x,y) = d(y,x)$ (*symmetry*), $d(x,z) \le d(x,y) + d(y,z)$ (*triangle inequality*). If additionally $d(x,y) = 0$ implies $x = y$, then d is a *metric*. A *(pseudo)metric space* (X,d) consists of a set X and a (pseudo)metric d on X.

A map $f\colon X \to [0,1]$ is *non-expansive* wrt. a pseudometric d if $|f(x) - f(y)| \le d(x,y)$ for all $x,y \in X$. The space of these non-expansive functions, denoted $\mathrm{Pred}(X,d)$, is equipped with the *supremum (pseudo)metric* d_∞,

$$d_\infty(f,g) = \|f-g\|_\infty = \textstyle\bigvee_{x\in X} |f(x) - g(x)|.$$

We denote by $B_\epsilon(x) = \{y \in X \mid d(x,y) \le \epsilon\}$ the *ball* of radius ϵ around x in (X,d). The space (X,d) is *totally bounded* if for every $\epsilon > 0$ there exists a finite ϵ*-cover*, i.e. finitely many elements $x_1,\dots,x_n \in X$ such that $X = \bigcup_{i=1}^n B_\epsilon(x_i)$. Recall that a metric space is compact iff it is complete and totally bounded.

We next introduce the Wasserstein and Kantorovich distances, which coincide according to Kantorovich-Rubinstein duality. To this end, we first need the notion of a coupling of two probability distributions, from which the original distributions are factored out as marginals.

Definition 3.2. Let π_1 and π_2 be discrete probability measures on A and B, respectively. We denote by $\mathrm{Cpl}(\pi_1,\pi_2)$ the set of *couplings* of π_1 and π_2, i.e. probability measures μ on $A \times B$ such that π_1 and π_2 are *marginals* of μ:

- for all $a \in A$, $\sum_{b\in B} \mu(a,b) = \pi_1(a)$;
- for all $b \in B$, $\sum_{a\in A} \mu(a,b) = \pi_2(b)$.

Definition 3.3 (Wasserstein and Kantorovich distances)**.** Let (X,d) be a pseudometric space. We generally write

$$\mathcal{D}X$$

for the set of discrete probability distributions on X. We define two pseudometrics on $\mathcal{D}X$, the *Kantorovich distance* $d^\uparrow$ and the *Wasserstein distance* $d^\downarrow$:

$$d^\uparrow(\pi_1,\pi_2) = \textstyle\bigvee\{|\,\mathrm{E}_{\pi_1}(f) - \mathrm{E}_{\pi_2}(f)| \mid f \in \mathrm{Pred}(X,d)\}$$
$$d^\downarrow(\pi_1,\pi_2) = \textstyle\bigwedge\{\mathrm{E}_\mu(d) \mid \mu \in \mathrm{Cpl}(\pi_1,\pi_2)\}$$

where $\bigwedge$ takes meets (and $\bigvee$ suprema). We extend these distances without further mention to zero functions (like the functions r_a at blocking states a) by decreeing that the zero function has distance 1 from all probability distributions.

The notation $d^\uparrow, d^\downarrow$ is meant as a mnemonic for the fact that these distances are obtained via suprema respectively via infima. If (X,d) is separable (contains a countable dense subset), these pseudometrics coincide, a fact known as the *Kantorovich-Rubinstein duality* (e.g. [Dudley, 2002]):

Lemma 3.4 (Kantorovich-Rubinstein duality)**.** *Let (X,d) be a separable pseudometric space. Then for all $\pi_1,\pi_2 \in \mathcal{D}X$,*

$$d^\uparrow(\pi_1,\pi_2) = d^\downarrow(\pi_1,\pi_2).$$

The above notions of *lifting* a distance on X to a distance on distributions over X can be used to give fixed point equations for behavioural distances on models.

Definition 3.5 (Fixed point iteration à la Wasserstein/Kantorovich)**.** Given a model $\mathcal{I}$, we define the chains (d_n^K), (d_n^W) of *depth-n Kantorovich* and *Wasserstein distances*, respectively, via fixed point iteration:

$$d_0^W(a,b) = d_0^K(a,b) = 0$$
$$d_{n+1}^W(a,b) = \textstyle\bigvee_{A\in\mathsf{N_C}} |A^{\mathcal{I}}(a) - A^{\mathcal{I}}(b)| \vee (d_n^W)^\downarrow(\pi_a,\pi_b)$$
$$d_{n+1}^K(a,b) = \textstyle\bigvee_{A\in\mathsf{N_C}} |A^{\mathcal{I}}(a) - A^{\mathcal{I}}(b)| \vee (d_n^K)^\uparrow(\pi_a,\pi_b)$$

where $\vee$ is binary join. We extend this to states a,b in different models $\mathcal{I}$, $\mathcal{J}$ by taking the disjoint union of $\mathcal{I}$, $\mathcal{J}$.

In both cases, we start with the zero pseudometric, and in the next iteration lift the pseudometric d_n from the previous step via Wasserstein/Kantorovich. This lifted metric is then applied to the probability distributions π_a, π_b associated with a, b. In addition we take the maximum with the supremum over the distances for all atomic $A \in \mathsf{N_C}$.

We now introduce a key tool for our technical development, a new up-to-ϵ bisimulation game inspired by the definition of the Wasserstein distance.

Definition 3.6 (Bisimulation game)**.** Given models $\mathcal{I}, \mathcal{J}$, $a_0 \in \Delta^{\mathcal{I}}, b_0 \in \Delta^{\mathcal{J}}$, and $\epsilon_0 \in [0,1]$, the *ϵ_0-bisimulation game* for a_0 and b_0 is played by *Spoiler* (S) and *Duplicator* (D), with rules as follows:

- *Configurations*: triples (a, b, ϵ), with states $a \in \Delta^{\mathcal{I}}$, $b \in \Delta^{\mathcal{J}}$ and maximal allowed deviation $\epsilon \in [0,1]$. The *initial configuration* is (a_0, b_0, ϵ_0).
- *Moves*: In each round, D first picks a probability measure $\mu \in \mathrm{Cpl}(\pi_a, \pi_b)$. Then, D distributes the deviation ϵ over all pairs (a', b') of successors, i.e. picks a function $\epsilon' \colon \Delta^{\mathcal{I}} \times \Delta^{\mathcal{J}} \to [0,1]$ such that $\mathrm{E}_\mu(\epsilon') \le \epsilon$. Finally, S picks a pair (a', b') with $\mu(a', b') > 0$; the new configuration is then $(a', b', \epsilon'(a', b'))$.
- *D wins* if both states are blocking or $\epsilon = 1$.
- *S wins* if exactly one state is blocking and $\epsilon < 1$.
- *Winning condition*: $|A^{\mathcal{I}}(a) - A^{\mathcal{J}}(b)| \le \epsilon$ for all $A \in \mathsf{N_C}$.

The game comes in two variants, the *(unbounded) bisimulation game* and the *n-round bisimulation game*, where $n \ge 0$. Player D wins if the winning condition holds *before* every round, otherwise S wins. More precisely, D wins the unbounded game if she can force an infinite play and the n-round game once n rounds have been played (the winning condition is not checked after the last round, so in particular, any 0-round game is an immediate win for D).

Remark 3.7. The above bisimulation game differs from bisimulation games in the literature (e.g. [Desharnais *et al.*, 2008]) in a number of salient features. A particularly striking aspect is that D's moves are not similar to those of S, and moreover D in fact moves before S. Intuitively, D is required to commit beforehand to a strategy that she will use to respond to S's next move. Note also that the precision ϵ changes as the game is being played, a complication forced by the arithmetic nature of models.

This leads to notions of game distance:

Definition 3.8. *depth-n game distance* d_n^G and (unbounded-depth) *game distance* d^G are defined as

$$d_n^G(a,b) = \bigwedge\{\epsilon \mid D \text{ wins } \mathsf{G}_n(a,b,\epsilon)\}$$
$$d^G(a,b) = \bigwedge\{\epsilon \mid D \text{ wins } \mathsf{G}(a,b,\epsilon)\}.$$

where $\mathsf{G}(a,b,\epsilon)$ and $\mathsf{G}_n(a,b,\epsilon)$ denote the the bisimulation game and the n-round bisimulation game on (a,b,ϵ), respectively.

Finally we define the depth-n logical distance via $\mathcal{ALC}(\mathsf{P})$, restricting to concepts of rank at most n:

Definition 3.9. The *depth-n logical distance* $d_n^L(a,b)$ of states a, b in models $\mathcal{I}, \mathcal{J}$ is defined as

$$d_n^L(a,b) = \bigvee\{|C^{\mathcal{I}}(a) - C^{\mathcal{J}}(b)| \mid \mathsf{rk}(C) \le n\}.$$

The equivalence of the four bounded-depth behavioural distances introduced above will be shown in Theorem 4.3.

Behavioural distance forms the yardstick for our notion of bisimulation invariance; for definiteness:

Definition 3.10. A quantitative, i.e. $[0,1]$-valued, property Q of states, or a formula or concept defining such a property, is *bisimulation-invariant* if Q is non-expansive wrt. game distance, i.e. for states a, b in models $\mathcal{I}, \mathcal{J}$, respectively,

$$|Q(a) - Q(b)| \le d^G(a,b).$$

Similarly, Q is *depth-n bisimulation invariant*, or *finite-depth bisimulation invariant* if mention of n is omitted, if Q is non-expansive wrt. d_n^G in the same sense.

It is easy to see that *$\mathcal{ALC}(\mathsf{P})$-concepts are bisimulation-invariant*. More precisely, $\mathcal{ALC}(\mathsf{P})$-concepts of rank at most n are depth-n bisimulation invariant (a stronger invariance since clearly $d_n^G \le d^G$), as shown by routine induction. In contrast, many other properties of states are expressible in $\mathsf{FO}(\mathsf{P})$ but not in $\mathcal{ALC}(\mathsf{P})$, as they fail to be bisimulation-invariant. Examples include $x\mathsf{P}\lceil y : x = y\rceil$ (probability of a self-transition) and $\exists z.\, x\mathsf{P}\lceil y : y = z\rceil$ (highest transition probability to a successor).

We are now ready to formally state our main theorem (a proof will be given in Section 6):

Theorem 3.11 (Modal characterization)**.** *Every bisimulation-invariant $\mathsf{FO}(\mathsf{P})$-formula of rank at most n can be approximated (uniformly across all models) by $\mathcal{ALC}(\mathsf{P})$-concepts of rank at most 3^n.*

(The exponential bound on the rank features also in the full statement of van Benthem's theorem.)

4 Modal Approximation at Finite Depth

We now establish the most important stepping stone on the way to the eventual proof of the modal characterization theorem: We show that every depth-n bisimulation-invariant property of states can be approximated by $\mathcal{ALC}(\mathsf{P})$-concepts of rank at most n. We prove this simultaneously with coincidence of the various finite-depth behavioural pseudometrics defined in the previous section. To begin,

Lemma 4.1. *The game-based pseudometric d_n^G coincides with the Wasserstein pseudometric d_n^W,*

We note next that the modality P is non-expansive: We extend P to act on $[0,1]$-valued functions $f \colon \Delta^{\mathcal{I}} \to [0,1]$ by

$$(\mathsf{P}f)(a) = \mathrm{E}_{r_a}(f).$$

Lemma 4.2. *The map $f \mapsto \mathsf{P}f$ is non-expansive wrt. the supremum metric, that is $\|\mathsf{P}f - \mathsf{P}g\|_\infty \le \|f - g\|_\infty$ for all $f, g \colon \Delta^{\mathcal{I}} \to [0,1]$.*

Following our previous work [Wild *et al.*, 2018], we prove coincidence of the remaining pseudometrics in one big induction, along with total boundedness (needed later to apply a variant of the Arzelà-Ascoli theorem and the Kantorovich-Rubinstein duality) and modal approximability of depth-n bisimulation-invariant properties. We phrase the latter as density of the (semantics of) $\mathcal{ALC}(\mathsf{P})$-concepts of rank at most n in the non-expansive function space (Definition 3.1):

Theorem 4.3. *Let $\mathcal{I}$ be a model. Then for all $n \geq 0$,*

1. *we have $d_n^G = d_n^W = d_n^K = d_n^L =: d_n$ on $\mathcal{I}$;*
2. *the pseudometric space $(\Delta^\mathcal{I}, d_n)$ is totally bounded;*
3. *$\mathcal{ALC}(\mathsf{P})_n$ is a dense subset of $\mathrm{Pred}(\Delta^\mathcal{I}, d_n)$.*

Proof sketch. By simultaneous induction on n.

In the base case $n = 0$, all the behavioural distances are the zero pseudometric, so that total boundedness follows trivially and the density claim follows because non-expansive maps are just constants in $[0, 1]$ and the syntax of $\mathcal{ALC}(\mathsf{P})$ includes truth constants $q \in \mathbb{Q} \cap [0, 1]$.

For the inductive step, let $\mathcal{I}$ be a model and $n > 0$, and assume as the inductive hypothesis that all claims in Theorem 4.3 hold for all $n' < n$. We begin with Item 1; $d_n^G = d_n^W$ is already proved (Lemma 4.1).

• $d_n^W = d_n^K$ follows by Kantorovich-Rubinstein duality (Lemma 3.4), since every totally bounded pseudometric space is separable.

• $d_n^K = d_n^L$: By Lemma 4.2 and the inductive hypothesis, $\mathsf{P}[\mathcal{ALC}(\mathsf{P})_{n-1}]$ is dense in $\mathsf{P}[\mathrm{Pred}(\Delta^\mathcal{I}, d_{n-1})]$. Thus, the supremum in the definition of d_n^K does not change when it is taken only over the concepts in $\mathcal{ALC}(\mathsf{P})_{n-1}$ instead of all nonexpansive properties. The proof is finished by a simple induction over propositional combinations of concepts.

Item 2: By the inductive hypothesis, the space $(\Delta^\mathcal{I}, d_{n-1})$ is totally bounded. By the Arzelà-Ascoli theorem (in a version for totally bounded spaces and non-expansive maps, cf. [Wild *et al.*, 2018]), it follows that $\mathrm{Pred}(\Delta^\mathcal{I}, d_{n-1})$ is totally bounded wrt. the supremum pseudometric. This implies that depth-n distances can be approximated up to ϵ by examining differences at only finitely many, say m, concepts. As $([0, 1]^m, d_\infty)$ is totally bounded, $(\Delta^\mathcal{I}, d_n)$ is, too.

Item 3: By the Stone-Weierstraß theorem (again in a version for totally bounded spaces and non-expansive maps [Wild *et al.*, 2018]) it suffices to give, for each $\epsilon > 0$, each non-expansive map $f \in \mathrm{Pred}(\Delta^\mathcal{I}, d_n)$, and each pair of states $a, b \in \Delta^\mathcal{I}$ a concept $C \in \mathcal{ALC}(\mathsf{P})_n$ such that

$$\max(|f(a) - C^\mathcal{I}(a)|, |f(b) - C^\mathcal{I}(b)|) \leq \epsilon.$$

To construct such a C, we note that $|f(a) - f(b)| \leq d_n^L(a, b)$ (by non-expansiveness), so there exists some $D \in \mathcal{ALC}(\mathsf{P})_n$ such that $|D^\mathcal{I}(a) - D^\mathcal{I}(b)| \geq |f(a) - f(b)| - \epsilon$. From D, we can construct C using truncated subtraction $\ominus$. □

This completes the proof of Theorem 4.3. Now that we can approximate depth-k bisimulation-invariant properties by $\mathcal{ALC}(\mathsf{P})$-concepts of rank k on any fixed model, we need to make the approximation uniform across all models. We achieve this by means of a *final* model, i.e. one that realizes all behaviours. Formally:

Definition 4.4. A *(probabilistic) bounded morphism* between models $\mathcal{I}$, $\mathcal{J}$ is a map $f : \Delta^\mathcal{I} \to \Delta^\mathcal{J}$ such that $A^\mathcal{I} = f^{-1}[A^\mathcal{J}]$ for each $A \in \mathsf{N_C}$ and $r_{f(a)}(B) = r_a(f^{-1}[B])$ for all $B \subseteq \Delta^\mathcal{J}$, $a \in \Delta^\mathcal{I}$ (implying that a is blocking iff $f(a)$ is blocking). A model $\mathcal{F}$ is *final* if for every model $\mathcal{I}$, there exists a unique bounded morphism $\mathcal{I} \to \mathcal{F}$.

It follows from standard results in coalgebra [Barr, 1993] that a final model exists. Bounded morphisms preserve behaviour on-the-nose, that is:

Lemma 4.5. *Let $f : \mathcal{I} \to \mathcal{J}$ be a bounded morphism. Then, for any $a \in \Delta^\mathcal{I}$, $d^G(a, f(a)) = 0$.*

This entails the following lemma, which will enable us to use approximants on the final model as uniform approximants across all models:

Lemma 4.6. *Let $\mathcal{F}$ be a final model, and let ϕ and ψ be bisimulation-invariant first-order properties. Then, for any model $\mathcal{I}$, $\|\phi - \psi\|_\infty^\mathcal{I} \leq \|\phi - \psi\|_\infty^\mathcal{F}$.*

5 Locality

The proof of the modal characterization theorem now further proceeds by first establishing that every bisimulation-invariant first-order formula ϕ is *local* in a sense to be made precise shortly, and subsequently that ϕ is in fact even finite-depth bisimulation invariant, for a depth that is exponential in the rank of ϕ. Locality refers to a probabilistic variant of Gaifman graphs [Gaifman, 1982]:

Definition 5.1. Let $\mathcal{I}$ be a model.

• The *Gaifman graph* of $\mathcal{I}$ is the undirected graph on the set $\Delta^\mathcal{I}$ of vertices that has an edge for every pair (a, b) with $r^\mathcal{I}(a, b) > 0$ or $r^\mathcal{I}(b, a) > 0$.

• The *Gaifman distance* $D\colon \Delta^\mathcal{I} \times \Delta^\mathcal{I} \to \mathbb{N} \cup \{\infty\}$ is graph distance in the Gaifman graph: For every $a, b \in \Delta^\mathcal{I}$, the distance $D(a, b)$ is the least number of edges on any path from a to b, if such a path exists, and ∞ otherwise.

• For $a \in \Delta^\mathcal{I}$ and $k \geq 0$, the *radius k neighbourhood* $U^k(a) = \{b \in \Delta^\mathcal{I} \mid D(a, b) \leq k\}$ of a consists of the states reachable from a in at most k steps.

• The *restriction* of $\mathcal{I}$ to $U^k(a)$ is the model $\mathcal{I}_a^k$ with set $U^k(a)$ of states, and

$$A^{\mathcal{I}_a^k}(b) = A^\mathcal{I}(b) \quad r^{\mathcal{I}_a^k}(b, c) = \begin{cases} r^\mathcal{I}(b, c) & \text{if } D(a, b) < k \\ 0 & \text{if } D(a, b) = k \end{cases}$$

for $A \in \mathsf{N_C}$ and $b, c \in U^k(a)$.

The restriction to $U^k(a)$ thus makes all states at distance k blocking. Restricted models have the expected relationship with games of bounded depth:

Lemma 5.2. *Let a be a state in a model $\mathcal{I}$. Then D wins the k-round 0-bisimulation game for $\mathcal{I}, a$ and $\mathcal{I}_a^k, a$.*

Locality of a formula now means that its truth values only depend on the neighbourhood of the state in question:

Definition 5.3. A formula $\phi(x)$ is *k-local* for a radius k if for every model $\mathcal{I}$ and every $a \in \Delta^\mathcal{I}$, $\phi^\mathcal{I}(a) = \phi^{\mathcal{I}_a^k}(a)$.

As $\mathcal{ALC}(\mathsf{P})$-concepts are bisimulation-invariant, Lemma 5.2 implies

Lemma 5.4. *Every $\mathcal{ALC}(\mathsf{P})$-concept of rank at most k is k-local.*

To prove locality of bisimulation-invariant $\mathsf{FO}(\mathsf{P})$-formulae, we require a model-theoretic tool, an adaptation of Ehrenfeucht-Fraïssé equivalence to the probabilistic setting:

Definition 5.5. Let $\mathcal{I}, \mathcal{J}$ be models, and let $\bar{a}_0$ and $\bar{b}_0$ be vectors of equal length over $\Delta^\mathcal{I}$ and $\Delta^\mathcal{J}$, respectively. The *Ehrenfeucht-Fraïssé game for* $\mathcal{I}, \bar{a}_0$ *and* $\mathcal{J}, \bar{b}_0$, played by *Spoiler* (S) and *Duplicator* (D), is given as follows.

- *Configurations:* pairs $(\bar{a}, \bar{b})$ of vectors $\bar{a}$ over $\Delta^\mathcal{I}$ and $\bar{b}$ over $\Delta^\mathcal{J}$; the *initial configuration* is $(\bar{a}_0, \bar{b}_0)$.
- *Moves:* Each round can be played in one of two ways, chosen by S:

  – *Standard round*: S selects a state in one model, say $a \in \Delta^\mathcal{I}$, and D then has to select a state in the other model, say $b \in \Delta^\mathcal{J}$, reaching the configuration $(\bar{a}a, \bar{b}b)$.

  – *Probabilistic round*: S selects an index i and a fuzzy subset in one model, say $\phi_A \colon \Delta^\mathcal{I} \to [0, 1]$. D then has to select a fuzzy subset in the other model, say $\phi_B \colon \Delta^\mathcal{J} \to [0, 1]$, such that $\mathrm{E}_{r_{a_i}}(\phi_A) = \mathrm{E}_{r_{b_i}}(\phi_B)$. Then, S selects an element on one side, say $a \in \Delta^\mathcal{I}$, such that $r_{a_i}(a) > 0$, and D subsequently selects an element on the other side, say $b \in \Delta^\mathcal{J}$, such that $\phi_A(a) = \phi_B(b)$ and $r_{b_i}(b) > 0$, reaching the configuration $(\bar{a}a, \bar{b}b)$.
- *Winning conditions*: Any player who cannot move loses. S wins if a configuration is reached (including the initial configuration) that fails to be a partial isomorphism. Here, a configuration $(\bar{a}, \bar{b})$ is a *partial isomorphism* if
  - $a_i = a_j \iff b_i = b_j$
  - $A^\mathcal{I}(a_i) = A^\mathcal{J}(b_i)$ for all i and all $A \in \mathsf{N_C}$
  - $r^\mathcal{I}(a_i, a_j) = r^\mathcal{J}(b_i, b_j)$ for all i, j.

  Player D wins if she reaches the n-th round (maintaining configurations that are not winning for S).

For our purposes, we need only soundness of Ehrenfeucht-Fraïssé equivalence:

Lemma 5.6 (Ehrenfeucht-Fraïssé invariance)**.** *Let* $\mathcal{I}, \mathcal{J}$ *be models, and let* $\bar{a}_0, \bar{b}_0$ *be vectors of length* m *over* $\Delta^\mathcal{I}$ *and* $\Delta^\mathcal{J}$*, respectively, such that* D *wins the* n*-round Ehrenfeucht-Fraïssé game on* $\bar{a}_0, \bar{b}_0$*. Then for every* $\mathsf{FO}(\mathit{P})$*-formula* ϕ *with* $\mathsf{qr}(\phi) \leq n$ *and free variables at most* $x_1, \dots, x_m$*,*

$$\phi(\bar{a}_0) = \phi(\bar{b}_0).$$

Since embeddings into disjoint unions of models are bounded morphisms, the following is immediate from Lemma 4.5:

Lemma 5.7. *Every bisimulation-invariant formula is also invariant under disjoint union.*

We are now in a position to prove our desired locality result:

Lemma 5.8 (Locality)**.** *Let* $\phi(x)$ *be a bisimulation-invariant* $\mathsf{FO}(\mathit{P})$*-formula of rank* n *with one free variable* x*. Then* ϕ *is* k*-local for* $k = 3^n$*.*

Proof sketch. Let a be a state in a model $\mathcal{I}$. We need to show $\phi^\mathcal{I}(a) = \phi^{\mathcal{I}^k_a}(a)$. Construct models $\mathcal{J}, \mathcal{K}$ that extend $\mathcal{I}$ and $\mathcal{I}^k_a$, respectively, by adding n disjoint copies of both $\mathcal{I}$ and $\mathcal{I}^k_a$. We finish the proof by showing that

$$\phi^\mathcal{I}(a) = \phi^\mathcal{J}(a) = \phi^\mathcal{K}(a) = \phi^{\mathcal{I}^k_a}(a).$$

The first and third equality follow by bisimulation invariance of ϕ (Lemma 5.7), and the second using Lemma 5.6, by giving a winning invariant for D in the n-round Ehrenfeucht-Fraïssé game for $\mathcal{J}, a$ and $\mathcal{K}, a$. □

6 Proof of the Main Result

Having established locality of bisimulation-invariant first-order formulae and modal approximability of finite-depth bisimulation-invariant properties, we now discharge the last remaining steps in our programme: We show by means of an unravelling construction that bisimulation-invariant first-order formulae are already finite-depth bisimulation-invariant, and then conclude the proof of our main result, the modal characterization theorem.

Definition 6.1. Let $\mathcal{I}$ be a model. The *unravelling* $\mathcal{I}^*$ of $\mathcal{I}$ is a model with non-empty finite sequences $\bar{a} \in (\Delta^\mathcal{I})^+$ as states, where atomic concepts and roles are interpreted by

$$A^{\mathcal{I}^*}(\bar{a}) = A^\mathcal{I}(\mathsf{last}(\bar{a})) \qquad r^{\mathcal{I}^*}(\bar{a}, \bar{a}a) = r^\mathcal{I}(\mathsf{last}(\bar{a}), a),$$

for $\bar{a} \in (\Delta^\mathcal{I})^+$ and $a \in \Delta^\mathcal{I}$, where last takes last elements.

As usual, models are bisimilar to their unravellings:

Lemma 6.2. *For any model* $\mathcal{I}$ *and* $a \in \Delta^\mathcal{I}$*,* D *has a winning strategy in the* 0*-bisimulation game for* $\mathcal{I}, a$ *and* $\mathcal{I}^*, a$*.*

We next show that locality and bisimulation invariance imply finite-depth bisimulation invariance:

Lemma 6.3. *Let* ϕ *be bisimulation invariant and* k*-local. Then* ϕ *is depth-*k *bisimulation invariant.*

Proof sketch. By unravelling (Lemma 6.2) and locality (Lemma 5.2), we need only consider depth-k tree models. On such models, winning strategies in k-round bisimulation games automatically win also the unrestricted game. □

This allows us to wrap up the proof of our main result:

Proof of Theorem 3.11. Let ϕ be a probabilistic first-order formula of rank n. By Lemma 5.8 and Lemma 6.3, ϕ is depth-k bisimulation-invariant for $k = 3^n$. By Theorem 4.3, for every $\epsilon > 0$, there exists an $\mathcal{ALC}(\mathsf{P})$ concept C_ϵ of rank at most k such that $\|\phi^\mathcal{F} - C_\epsilon^\mathcal{F}\|_\infty \leq \epsilon$ on the final model $\mathcal{F}$. By Lemma 4.6, this approximation works over all models. □

7 Conclusions

We have established a modal characterization result for a probabilistic fuzzy DL $\mathcal{ALC}(\mathsf{P})$, stating that every formula of quantitative probabilistic FOL that is *bisimulation-invariant*, i.e. non-expansive wrt. a natural notion of behavioural distance, can be approximated by $\mathcal{ALC}(\mathsf{P})$-concepts of bounded modal rank, the bound being exponential in the rank of the original formula. As discussed in the introduction, the bound on the modal rank is the crucial feature making this result into a van-Benthem (rather than Hennessy-Milner) type theorem.

It remains open whether our main result can be sharpened to make do without approximation. (Similar open problems persist for the case of fuzzy modal logic [Wild *et al.*, 2018] and two-valued probabilistic modal logic [Schröder *et al.*, 2017].) Further directions for future research include a treatment of Łukasiewicz semantics of the propositional connectives (for which non-expansiveness in fact fails). Moreover, the version of our main result that restricts the semantics to finite models, in analogy to Rosen's finite-model version of van Benthem's theorem [Rosen, 1997], remains open.

References

[Baldan *et al.*, 2014] P. Baldan, F. Bonchi, H. Kerstan, and B. König. Behavioral metrics via functor lifting. In *Found. Software Technology and Theoretical Computer Science, FSTTCS 2014*, *LIPIcs*, vol. 29, pp. 403–415, 2014.

[Barr, 1993] M. Barr. Terminal coalgebras in well-founded set theory. *Theor. Comput. Sci.*, 114:299–315, 1993.

[Burgess, 1969] J. Burgess. Probability logic. *J. Symb. Log.*, 34:264–274, 1969.

[Carreiro, 2015] F. Carreiro. PDL is the bisimulation-invariant fragment of weak chain logic. In *Logic in Computer Science, LICS 2015*, pp. 341–352. IEEE, 2015.

[Chang, 1973] C. Chang. Modal model theory. In *Cambridge Summer School in Mathematical Logic*, *LNM*, vol. 337, pp. 599–617. Springer, 1973.

[Dawar and Otto, 2005] A. Dawar and M. Otto. Modal characterisation theorems over special classes of frames. In *Logic in Comput. Sci., LICS 2005*, pp. 21–30. IEEE, 2005.

[Desharnais *et al.*, 1999] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov systems. In *Concurrency Theory, CONCUR 1999*, *LNCS*, vol. 1664, pp. 258–273. Springer, 1999.

[Desharnais *et al.*, 2004] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labelled Markov processes. *Theor. Comput. Sci.*, 318:323–354, 2004.

[Desharnais *et al.*, 2008] J. Desharnais, F. Laviolette, and M. Tracol. Approximate analysis of probabilistic processes: Logic, simulation and games. In *Quantitative Evaluation of Systems, QEST 2008*, pp. 264–273. IEEE, 2008.

[Dudley, 2002] R. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2nd edition, 2002.

[Figueira *et al.*, 2015] D. Figueira, S. Figueira, and C. Areces. Model theory of XPath on data trees. Part I: Bisimulation and characterization. *J. Artif. Intell. Res. (JAIR)*, 53:271–314, 2015.

[Flaminio and Godo, 2007] T. Flaminio and L. Godo. A logic for reasoning about the probability of fuzzy events. *Fuzzy Sets and Systems*, 158(6):625–638, 2007.

[Gaifman, 1982] H. Gaifman. On local and non-local properties. In *Logic Colloq. 1981*, pp. 105–135, 1982.

[Giacalone *et al.*, 1990] A. Giacalone, C. Jou, and S. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Programming Concepts and Methods*, pp. 443–458. North-Holland, 1990.

[Hájek, 2007] P. Hájek. Complexity of fuzzy probability logics II. *Fuzzy Sets and Systems*, 158:2605–2611, 2007.

[Halpern and Rabin, 1987] J. Halpern and M. Rabin. A logic to reason about likelihood. *Artif. Intell.*, 32:379–405, 1987.

[Halpern, 1990] J. Halpern. An analysis of first-order logics of probability. *Artif. Intell.*, 46:311–350, 1990.

[Hamblin, 1959] C. Hamblin. The modal 'probably'. *Mind*, 68:234–240, 1959.

[Heifetz and Mongin, 2001] A. Heifetz and P. Mongin. Probabilistic logic for type spaces. *Games Econ. Behav.*, 35:31–53, 2001.

[Hennessy and Milner, 1985] M. Hennessy and R. Milner. Algebraic laws for non-determinism and concurrency. *J. ACM*, 32:137–161, 1985.

[Herzig, 2003] A. Herzig. Modal probability, belief, and actions. *Fund. Inf.*, 57:323–344, 2003.

[Huth and Kwiatkowska, 1997] M. Huth and M. Kwiatkowska. Quantitative analysis and model checking. In *Logic in Computer Science, LICS 1997*, pp. 111–122. IEEE, 1997.

[Janin and Walukiewicz, 1995] D. Janin and I. Walukiewicz. Automata for the modal μ-calculus and related results. In *Mathematical Foundations of Computer Science, MFCS 1995*, *LNCS*, vol. 969, pp. 552–562. Springer, 1995.

[Kozen, 1985] D. Kozen. A probabilistic PDL. *J. Comput. Sys. Sci.*, 30(2):162–178, 1985.

[Larsen and Skou, 1991] K. Larsen and A. Skou. Bisimulation through probabilistic testing. *Inf. Comput.*, 94:1–28, 1991.

[Litak *et al.*, 2018] T. Litak, D. Pattinson, K. Sano, and L. Schröder. Model theory and proof theory of coalgebraic predicate logic. *Log. Meth. Comput. Sci.*, 14(1), 2018.

[Lukasiewicz and Straccia, 2008] T. Lukasiewicz and U. Straccia. Managing uncertainty and vagueness in description logics for the semantic web. *J. Web Sem.*, 6:291–308, 2008.

[Lutz and Schröder, 2010] C. Lutz and L. Schröder. Probabilistic description logics for subjective uncertainty. In *Principles of Knowledge Representation and Reasoning, KR 2010*. AAAI, 2010.

[Makowsky and Ziegler, 1980] J. Makowsky and M. Ziegler. Topological model theory with an interior operator: consistency properties and back - and forth arguments. *Arch. math. Logik*, 20:27–54, 1980.

[Otto, 2004] M. Otto. Elementary proof of the van Benthem-Rosen characterisation theorem. Technical Report 2342, TU Darmstadt, 2004.

[Otto, 2006] M. Otto. Bisimulation invariance and finite models. In *Logic Colloq. 2002*, pp. 276–298. ASL, 2006.

[Rosen, 1997] E. Rosen. Modal logic over finite structures. *J. Logic Lang. Inform.*, 6(4):427–439, 1997.

[Rutten, 2000] J. Rutten. Universal coalgebra: A theory of systems. *Theor. Comput. Sci.*, 249:3–80, 2000.

[Schröder and Pattinson, 2011] L. Schröder and D. Pattinson. Description logics and fuzzy probability. In *IJCAI 2011*, pp. 1075–1081, 2011.

[Schröder *et al.*, 2017] L. Schröder, D. Pattinson, and T. Litak. A van Benthem/Rosen theorem for coalgebraic predicate logic. *J. Log. Comput.*, 27(3):749–773, 2017.

[Sturm and Wolter, 2001] H. Sturm and F. Wolter. First-order expressivity for S5-models: Modal vs. two-sorted languages. *J. Philos. Logic*, 30:571–591, 2001.

[van Benthem, 1976] J. van Benthem. *Modal Correspondence Theory*. PhD thesis, Univ. of Amsterdam, 1976.

[van Breugel and Worrell, 2005] F. van Breugel and J. Worrell. A behavioural pseudometric for probabilistic transition systems. *Theor. Comput. Sci.*, 331:115–142, 2005.

[Wild and Schröder, 2017] P. Wild and L. Schröder. A characterization theorem for a modal description logic. In *IJCAI 2017*, pp. 1304–1310, 2017.

[Wild *et al.*, 2018] P. Wild, L. Schröder, D. Pattinson, and B. König. A van Benthem theorem for fuzzy modal logic. In *Logic in Computer Science, LICS 2018*, pp. 909–918. ACM, 2018.

[Zadeh, 1968] L. Zadeh. Probability measures of fuzzy events. *J. Math. Anal. Appl.*, 23:421–427, 1968.

A Appendix

A.1 Coalgebraic Modelling

Universal coalgebra [Rutten, 2000] serves as a generic framework for modelling state-based systems, with the system type encapsulated as a set functor. Although we are only concerned with a concrete system type in the present paper, we do need coalgebraic methods to some degree. In particular, the requisite background on behavioural distances [van Breugel and Worrell, 2005; Baldan *et al.*, 2014] is largely based on coalgebraic techniques, and moreover we will need the final coalgebra at one point in the development. We require only basic definitions, which we recapitulate here and then instantiate to the case of our notion of model.

Recall first that a set functor $F : \mathsf{Set} \to \mathsf{Set}$ consists of an assignment of a set FX to every set X and a map $Ff : FX \to FY$ to every map $f : X \to Y$, preserving identities and composition. The core example of a functor for the present purposes is the *distribution functor* $\mathcal{D}$, which assigns to a set X the set $\mathcal{D}X$ of discrete probability measures on X, and to a map $f : X \to Y$ the map $\mathcal{D}f : \mathcal{D}X \to \mathcal{D}Y$ that takes image measures; explicitly, $\mathcal{D}f(\mu)$ is the image measure of μ along f, given by $\mathcal{D}f(\mu)(A) = \mu(f^{-1}[A])$. Functors can be combined by taking *products* and *sums*: Given set functors $F, G : \mathsf{Set} \to \mathsf{Set}$, the set functors $F \times G, F + G : \mathsf{Set} \to \mathsf{Set}$ are given by $(F \times G)X = FX \times GX$ and $(F + G)X = FX + GX$, respectively, with the evident action on maps in both cases; here, $+$ denotes disjoint union as usual. Every set C induces a *constant functor*, also denoted C and given by $CX = C$ and $Cf = \mathsf{id}_C$ for every set X and every map f. Moreover, the *identity functor* id is given by $\mathsf{id}\,X = X$ and $\mathsf{id}\,f = f$ for all sets X and all maps f.

An *F-coalgebra* (A, ξ) for a set functor F consists of a set X of *states* and a *transition map* $\xi : A \to FA$, thought of as assigning to each state $a \in A$ a structured collection $\xi(a)$ of successors. A $\mathcal{D}$-coalgebra (A, ξ), for instance, is just a Markov chain: its transition map $\xi : A \to \mathcal{D}A$ assigns to each state a distribution over successor states. Similarly, models in the sense defined above are coalgebras (A, ξ) for the set functor $[0,1]^{\mathsf{N_C}} \times (\mathcal{D} + 1)$: If $\xi(a) = (f, \pi)$, then $f : \mathsf{N_C} \to [0, 1]$ determines the truth values of the atomic concepts at the state a, and π is either a discrete probability measure determining the successors of a or a designated value denoting termination. The probabilistic transition systems considered by van Breugel and Worrell [van Breugel and Worrell, 2005], which indexes probabilistic transition relations over a set Act of actions and moreover uses unrestricted subdistributions, corresponds to coalgebras (A, ξ) for the set functor $\mathcal{D}(\mathsf{id} + 1)^{\mathsf{Act}}$ – given a state a and an action $c \in \mathsf{Act}$, $\xi(a)(c) \in \mathcal{D}(A + 1)$ is a subdistribution over successor states of a, with the summand 1 serving to absorb the weight missing to obtain total weight 1.

A *morphism* $f : (A, \xi) \to (B, \zeta)$ between F-coalgebras (A, ξ) and (B, ζ) is a map $f : A \to B$ such that

$$Ff(\xi(a)) = \zeta(f(a))$$

for all states $a \in A$. Morphisms should be thought of as behaviour-preserving maps or functional bisimulations. E.g. $f : A \to B$ is a morphism of $\mathcal{D}$-coalgebras (i.e. Markov chains) (A, ξ) and (B, ζ) if for each set $Y \subseteq B$ and each state $a \in A$,

$$\zeta(f(a))(Y) = \xi(a)(f^{-1}[Y]),$$

i.e. the probability of reaching Y from $f(a)$ is the same as that of reaching $f^{-1}[Y]$ from a. Morphisms of probabilistic transition systems, viewed as coalgebras, satisfy a similar condition for the successor distributions, and additionally preserve the truth values of atomic concepts.

An F-coalgebra (Z, ζ) is *final* if for every F-coalgebra (A, ξ) there exists exactly one morphism $(A, \xi) \to (Z, \zeta)$. Final coalgebras are unique up to isomorphism if they exist, and should be thought of as having as states all possible behaviours of states in F-coalgebras. For our present purposes, we do not need an explicit description of the final coalgebra; it suffices to know that since the functor describing probabilistic transition systems is *accessible* (more precisely ω_1-accessible), a final coalgebra for it, i.e. a final probabilistic transition system, exists [Barr, 1993].

A.2 Omitted Proofs

Proof of Lemma 3.4

We make use of the following version of the Kantorovich-Rubinstein duality [Dudley, 2002, Proposition 11.8.1]:

Lemma A.1 (Kantorovich-Rubinstein duality)**.** *Let (X, d) be a separable metric space, and let $\mathcal{P}_1(X)$ denote the space of probability measures $\mu\colon \mathcal{B}(X) \to [0, 1]$ on the Borel σ-algebra $\mathcal{B}(X)$ such that $\int d(x, \cdot\,)\,\mathrm{d}\mu < \infty$ for some $x \in X$. Then for $\mu_1, \mu_2 \in \mathcal{P}_1(X)$,*

$$d^{\uparrow}(\mu_1, \mu_2) = d^{\downarrow}(\mu_1, \mu_2).$$

Essentially, we only need to transfer this version of Kantorovich-Rubinstein duality to the slightly more general case of pseudometrics.

First, note that the relation $x \sim y :\iff d(x, y) = 0$ is an equivalence relation on X. The quotient set $Y := X/{\sim}$ is made into a metric space (Y, d'), the *metric quotient* of (X, d), by taking $d'([x], [y]) = d(x, y)$. Let $p\colon A \to B$ be the projection map. By construction, p is an isometry. Both the Kantorovich and the Wasserstein lifting preserve isometries [Baldan *et al.*, 2014], so for all discrete probability measures μ_1, μ_2 on X,

$$\begin{aligned} d^{\uparrow}(\mu_1, \mu_2) &= (d')^{\uparrow}((\mathcal{D}p)\mu_1, (\mathcal{D}p)\mu_2) \\ &= (d')^{\downarrow}((\mathcal{D}p)\mu_1, (\mathcal{D}p)\mu_2) \\ &= d^{\downarrow}(\mu_1, \mu_2). \end{aligned}$$

In the second step we have applied Lemma A.1 to the metric space (Y, d'), noting that every discrete probability measure can be defined on the Borel σ-algebra.

Proof of Lemma 4.1.

Induction over n. The base case $n = 0$ is clear: the 0-round game is an immediate win for D, so $d_0^G = d_0^W = 0$. We proceed with the inductive step from n to $n + 1$.

So let a and b be states in a model $\mathcal{I}$. If a and b are both blocking, then $d^G_{n+1}(a,b) = d^W_{n+1}(a,b) = 0$. If exactly one of a, b is blocking, then $d^G_{n+1}(a,b) = d^W_{n+1}(a,b) = 1$. Now assume that both a and b are non-blocking.

"$\geq$": Let $d^G_{n+1}(a,b) \leq \epsilon$, so D wins the $(n+1)$-round bisimulation game on (a,b,ϵ). We show that $d^W_{n+1}(a,b) \leq \epsilon$. First, for every $A \in \mathsf{N_C}$, $|A^{\mathcal{I}}(a) - A^{\mathcal{I}}(b)| \leq \epsilon$ by the winning condition. Second, suppose D chooses $\mu \in \mathrm{Cpl}(r_a, r_b)$ and $\epsilon' \colon \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \to [0,1]$ in the first turn. By assumption, D wins the n-round bisimulation game on $(a', b', \epsilon'(a',b'))$ for every $a', b' \in \Delta^{\mathcal{I}}$, so $d^W_n = d^G_n \leq \epsilon'$ by induction, and thus $\mathrm{E}_\mu(d^W_n) \leq \mathrm{E}_\mu(\epsilon') \leq \epsilon$.

"$\leq$": Let $d^W_{n+1}(a,b) < \epsilon$. It suffices to give a winning strategy for D in the $(n+1)$-round bisimulation game on (a,b,ϵ) (implying $d^G_{n+1}(a,b) \leq \epsilon$). The winning condition in the initial configuration follows immediately from the assumption. Also by the assumption, there exists $\mu \in \mathrm{Cpl}(r_a, r_b)$ such that $\mathrm{E}_\mu(d^W_n) < \epsilon$. As r_a and r_b are discrete, the set

$$R := \{(a', b') \mid r_a(a') > 0 \text{ and } r_b(b') > 0\}$$

is countable; so we can write $R = \{(a_1, b_1), (a_2, b_2), \dots\}$. Now put $\delta = \epsilon - \mathrm{E}_\mu(d^W_n)$ and define

$$\epsilon'(a_i, b_i) = d^W_n(a_i, b_i) + 2^{-i}\delta$$

for $(a_i, b_i) \in R$ and $\epsilon'(a', b') = 0$ for $(a', b') \notin R$. Then

$$\mathrm{E}_\mu(\epsilon') \leq \mathrm{E}_\mu(d^W_n) + \delta = \epsilon,$$

so playing μ and ϵ' constitutes a legal move for D. Now, since $\mu \in \mathrm{Cpl}(r_a, r_b)$, $\mu(a', b') = 0$ for all $(a', b') \notin R$. This means that S must pick some $(a_i, b_i) \in R$. Then

$$d^G_n(a_i, b_i) = d^W_n(a_i, b_i) < \epsilon'(a_i, b_i),$$

so D wins the n-round game on $(a_i, b_i, \epsilon'(a_i, b_i))$.

Proof of Lemma 4.2.

Let $\|f - g\|_\infty \leq \epsilon$; we have to show $\|\mathsf{P}f - \mathsf{P}g\|_\infty \leq \epsilon$. So let $a \in \Delta^{\mathcal{I}}$; then

$$|(\mathsf{P}f)(a) - (\mathsf{P}g)(a)| = \mathrm{E}_{r_a}(f - g) \leq \mathrm{E}_{r_a}(\epsilon) \leq \epsilon,$$

as required.

Proof of Theorem 4.3.

We proceed by simultaneous induction on n.

In the base case $n = 0$, all the behavioural distances are the zero pseudometric: $d^G_0 = 0$ because by the rules of the game each 0-round game is an immediate win for D; $d^W_0 = d^K_0 = 0$ by definition; and $d^L_0 = 0$ because each rank-0 concept is a propositional combination of truth constants and therefore constant. Total boundedness follows directly from the fact that under the zero pseudometric every ϵ-ball is the entire space, regardless of ϵ. Finally, the density claim follows because non-expansive maps under the zero pseudometric are just constants in $[0,1]$ and the syntax of $\mathcal{ALC}(\mathsf{P})$ includes truth constants $q \in \mathbb{Q} \cap [0,1]$.

For the inductive step, let $\mathcal{I}$ be a model and $n > 0$, and assume as the inductive hypothesis that all claims in Theorem 4.3 hold for all $n' < n$. We begin with Item 1:

- $d^G_n = d^W_n$ is Lemma 4.1.
- $d^W_n = d^K_n$ follows by Kantorovich-Rubinstein duality (Lemma 3.4), since every totally bounded pseudometric space is separable.
- $d^K_n = d^L_n$: Let $a, b \in \Delta^{\mathcal{I}}$ and consider the map

$$G \colon \mathrm{Pred}(\Delta^{\mathcal{I}}, d_{n-1}) \to [0,1], \quad f \mapsto |(\mathsf{P}f)(a) - (\mathsf{P}f)(b)|,$$

Then G is a continuous function because all of its constituents are continuous (in particular, P is continuous by Lemma 4.2).

By the induction hypothesis, and because density is preserved by continuous maps, $G[\mathcal{ALC}(\mathsf{P})_{n-1}]$ is a dense subset of $G[\mathrm{Pred}(\Delta^{\mathcal{I}}, d_{n-1})]$. Thus,

$$\begin{aligned}
&d^K_n(a,b) \\
&= \bigvee_{A \in \mathsf{N_C}} |A^{\mathcal{I}}(a) - A^{\mathcal{I}}(b)| \vee \bigvee G[\mathrm{Pred}(\Delta^{\mathcal{I}}, d_{n-1})] \\
&= \bigvee_{A \in \mathsf{N_C}} |A^{\mathcal{I}}(a) - A^{\mathcal{I}}(b)| \vee \bigvee G[\mathcal{ALC}(\mathsf{P})_{n-1}] \\
&= \bigvee_{A \in \mathsf{N_C}} |A^{\mathcal{I}}(a) - A^{\mathcal{I}}(b)| \vee \bigvee_{\mathsf{rk}\, C \leq n-1} |(\mathsf{P}C)^{\mathcal{I}}(a) - (\mathsf{P}C)^{\mathcal{I}}(b)| \\
&= \bigvee_{\mathsf{rk}\, C \leq n} |C^{\mathcal{I}}(a) - C^{\mathcal{I}}(b)| = d^L_n(a,b).
\end{aligned}$$

To prove the penultimate step, we first note that "$\leq$" follows immediately. To see "$\geq$", we proceed by induction over the propositional combinations of atomic concepts $A \in \mathsf{N_C}$ and concepts $\mathsf{P}C$, where $C \in \mathcal{ALC}(\mathsf{P})_{n-1}$, using that for any concepts C, D and $q \in \mathbb{Q} \cap [0,1]$:

$$\begin{aligned}
|(C \ominus q)^{\mathcal{I}}(a) - (C \ominus q)^{\mathcal{I}}(b)| &\leq |C^{\mathcal{I}}(a) - C^{\mathcal{I}}(b)| \\
|(\neg C)^{\mathcal{I}}(a) - (\neg C)^{\mathcal{I}}(b)| &= |C^{\mathcal{I}}(a) - C^{\mathcal{I}}(b)| \\
|(C \sqcap D)^{\mathcal{I}}(a) - (C \sqcap D)^{\mathcal{I}}(b)| & \\
&\leq \max(|C^{\mathcal{I}}(a) - C^{\mathcal{I}}(b)|, |D^{\mathcal{I}}(a) - D^{\mathcal{I}}(b)|).
\end{aligned}$$

Item 2: We make use of the following version of the Arzelà-Ascoli theorem [Wild *et al.*, 2018] where function spaces are restricted to non-expansive functions instead of the more general continuous functions, but the underlying spaces are only required to be totally bounded instead of compact:

Lemma A.2 (Arzelà-Ascoli for totally bounded spaces)**.** *Let (X, d) be a totally bounded pseudometric space. Then the space $\mathrm{Pred}(X, d)$, equipped with the supremum pseudometric, is totally bounded.*

By Lemma A.2, applied to the inductive hypothesis, we know that the space $\mathrm{Pred}(\Delta^{\mathcal{I}}, d_{n-1})$ is totally bounded wrt. the supremum pseudometric.

Let $\epsilon > 0$. As $\mathcal{ALC}(\mathsf{P})_{n-1}$ is dense in $\mathrm{Pred}(\Delta^{\mathcal{I}}, d_{n-1})$, there exist finitely many $C_1, \dots, C_m \in \mathcal{ALC}(\mathsf{P})_{n-1}$ such that

$$\bigcup_{i=1}^{m} B_{\frac{\epsilon}{8}}(C_i) = \mathrm{Pred}(\Delta^{\mathcal{I}}, d_{n-1})$$

From these concepts, together with the atomic concepts $A_1, \dots, A_k$, we can construct the map

$$\begin{aligned}
&I \colon \Delta^{\mathcal{I}} \to [0,1]^{k+m} \\
&\quad a \mapsto (A_1^{\mathcal{I}}(a), \dots, A_k^{\mathcal{I}}(a), (\mathsf{P}C_1)^{\mathcal{I}}(a), \dots, (\mathsf{P}C_m)^{\mathcal{I}}(a)).
\end{aligned}$$

Note that we assume here that the set of atomic concepts is a finite set $\mathsf{N_C} = \{A_1, \dots, A_k\}$. This is without loss of generality for the modal characterization theorem, because every formula of $\mathsf{FO}(\mathsf{P})$ can only contain finitely many propositional atoms, so $\mathsf{N_C}$ can be restricted to just those atoms.

It turns out that I is an $\frac{\epsilon}{4}$-isometry, that is

$$|d_n(a,b) - \|I(a) - I(b)\|_\infty| \le \tfrac{\epsilon}{4}$$

for all $a, b \in \Delta^\mathcal{I}$. Thus, by the triangle inequality, we can take preimages to turn a finite $\frac{\epsilon}{4}$-cover of $[0,1]^{k+m}$ (a compact, hence totally bounded space) into a finite ϵ-cover of $(\Delta^\mathcal{I}, d_n)$.

Item 3: We make use of the following Stone-Weierstraß theorem [Wild *et al.*, 2018] (again in a version for totally bounded spaces and non-expansive maps):

Lemma A.3 (Stone-Weierstraß for totally bounded spaces)**.** *Let (X, d) be a totally bounded pseudometric space, and let L be a subset of $\mathrm{Pred}(X, d)$ such that $f_1, f_2 \in L$ implies $\min(f_1, f_2), \max(f_1, f_2) \in L$. Then L is dense in $\mathrm{Pred}(X, d)$ if each $f \in \mathrm{Pred}(X, d)$ can be approximated at each pair of points by functions in L; that is for all $\epsilon > 0$ and all $x_1, x_2 \in X$ there exists $g \in L$ such that*

$$\max(|f(x_1) - g(x_1)|, |f(x_2) - g(x_2)|) \le \epsilon.$$

We apply Lemma A.3 to $(\Delta^\mathcal{I}, d_n)$ with $L := \mathcal{ALC}(\mathsf{P})_n$. Clearly L is closed under $\min$ and $\max$ so, to finish the proof, it suffices to give, for each $\epsilon > 0$, each non-expansive map $f \in \mathrm{Pred}(\Delta^\mathcal{I}, d_n)$ and each pair of states $a, b \in \Delta^\mathcal{I}$ a concept $C \in \mathcal{ALC}(\mathsf{P})_n$ such that

$$\max(|f(a) - C^\mathcal{I}(a)|, |f(b) - C^\mathcal{I}(b)|) \le \epsilon.$$

To construct such a C, we note that $|f(a) - f(b)| \le d_n^L(a,b)$ (by non-expansiveness), so there exists some $D \in \mathcal{ALC}(\mathsf{P})_n$ such that $|D^\mathcal{I}(a) - D^\mathcal{I}(b)| \ge |f(a) - f(b)| - \epsilon$. From D, we can construct C using truncated subtraction $\ominus$.

Proof of Lemma 4.5.

We show that D wins the bisimulation game for $(a_0, f(a_0), 0)$ by maintaining the invariant that the current configuration is of the form $(a, b, 0)$ with $b = f(a)$, which ensures that the winning condition always holds. It remains to show that D can maintain the invariant.

In each round, D begins by picking $\mu(a', b') = r_a(a')$ if $b' = f(a')$ and 0 otherwise, and $\epsilon' = 0$. We can see that $\mu \in \mathrm{Cpl}(r_a, r_b)$, because

$$\textstyle\sum_{b' \in \Delta^\mathcal{J}} \mu(a', b') = r_a(a')$$

and

$$\textstyle\sum_{a' \in \Delta^\mathcal{I}} \mu(a', b') = \sum_{f(a') = b'} r_a(a') = r_b(b')$$

for all $a' \in \Delta^\mathcal{I}$ and $b' \in \Delta^\mathcal{J}$. Also, clearly $\mathrm{E}_\mu(\epsilon') = 0$. Now any choice by S leads to another configuration $(a', b', 0)$ with $b' = f(a')$.

Proof of Lemma 4.6.

Let $\mathcal{I}$ be a model, and let $h \colon \mathcal{I} \to \mathcal{F}$ be the unique morphism. Let $a \in \Delta^\mathcal{I}$. Then $d^G(a, h(a)) = 0$ by Lemma 4.5, and thus $\phi^\mathcal{I}(a) = \phi^\mathcal{F}(h(a))$ and $\psi^\mathcal{I}(a) = \psi^\mathcal{F}(h(a))$ by bisimulation invariance. So

$$\begin{aligned}\|\phi - \psi\|_\infty^\mathcal{I} &= \textstyle\bigvee_{a \in \Delta^\mathcal{I}} |\phi^\mathcal{I}(a) - \psi^\mathcal{I}(a)| \\ &= \textstyle\bigvee_{a \in \Delta^\mathcal{I}} |\phi^\mathcal{F}(h(a)) - \psi^\mathcal{F}(h(a))| \\ &\le \|\phi - \psi\|_\infty^\mathcal{F}.\end{aligned}$$

Proof of Lemma 5.2.

Player D wins by maintaining the invariant that whenever i rounds have been played, the current configuration is of the form $(a_i, a_i, 0)$ for some $a_i \in \Delta^\mathcal{I}$ with $D(a, a_i) \le i$. For $i < k$, no configuration of this kind can be winning for S, because the two states in this configuration represent the same state in different models (recall that the winning conditions are not checked after the last round has been played).

It remains to give a strategy for D that maintains the invariant. It clearly holds at the start of the game, with $a_0 = a$. When the $(i+1)$-th round is played, D can pick $\mu \in \mathrm{Cpl}(r_{a_i}, r_{a_i})$ and $\epsilon' \colon \Delta^\mathcal{I} \times U^k(a) \to [0,1]$ as follows:

$$\begin{aligned}\mu(a', a'') &= \begin{cases}\pi_{a_i}(a'), & \text{if } a' = a'', \\ 0, & \text{otherwise,}\end{cases} \\ \epsilon'(a', a'') &= 0.\end{aligned}$$

Clearly, $\mathrm{E}_\mu(\epsilon') = 0$, so this is a legal move. Now the new configuration chosen by S necessarily satisfies the invariant.

Proof of Lemma 5.6.

We proceed by induction over formulae.

- The cases $A(x_i)$ and $x_i = x_j$ (with $A \in \mathsf{N_C}$) follow immediately from the fact that the initial configuration is a partial isomorphism.
- The Boolean cases ($q, \phi \ominus q, \neg\phi, \phi \sqcap \psi$) follow directly by the inductive hypothesis.
- $\exists x.\, \phi$: Let $(\bar{a}, \bar{b})$ be the current configuration. Let $\delta > 0$, let a be such that

  $$(\exists x.\, \phi)(\bar{a}) - \phi(\bar{a}a) < \delta,$$

  and let b be the winning answer for D in reply to S choosing a. By induction, $\phi(\bar{a}a) = \phi(\bar{b}b)$, so

  $$(\exists x.\, \phi)(\bar{b}) \ge \phi(\bar{b}b) = \phi(\bar{a}a) > (\exists x.\, \phi)(\bar{a}) - \delta.$$

  Because $\delta > 0$ was arbitrary, it follows that $(\exists x.\, \phi)(\bar{b}) \ge (\exists x.\, \phi)(\bar{a})$. We can symmetrically show that $(\exists x.\, \phi)(\bar{a}) \ge (\exists x.\, \phi)(\bar{b})$, which proves this case.
- $x_i \mathsf{P}\lceil x_{m+1} : \phi\rceil$: Let $(\bar{a}, \bar{b})$ be the current configuration. Suppose that S picks the index i and the fuzzy subset

  $$\phi_A \colon \Delta^\mathcal{I} \to [0,1], \quad a \mapsto \phi^\mathcal{I}(\bar{a}a)$$

  and D's winning reply is $\psi_B \colon \Delta^\mathcal{J} \to [0,1]$. We show that on the support of r_{b_i}, ψ_B must be equal to

  $$\phi_B \colon \Delta^\mathcal{J} \to [0,1], \quad b \mapsto \phi^\mathcal{J}(\bar{b}b).$$

  Suppose there exists some $b \in \Delta^\mathcal{J}$ with $r(b_i, b) > 0$ and $\phi_B(b) \ne \psi_B(b)$. Then D has a winning reply $a \in \Delta^\mathcal{I}$ in case S picks this b, which means, by the rules of the game, that $r(a_i, a) > 0$ and $\phi_A(a) = \psi_B(b)$. However, it is also true that $\phi_A(a) = \phi_B(b)$, by the inductive hypothesis. This is a contradiction.

  Now, because ψ_B was a winning reply, we obtain

  $$\begin{aligned}(x_i \mathsf{P}\lceil x_{m+1} : \phi\rceil)(\bar{a}) &= \mathrm{E}_{r_{a_i}}(\phi_A) \\ &= \mathrm{E}_{r_{b_i}}(\psi_B) \\ &= \mathrm{E}_{r_{b_i}}(\phi_B) \\ &= (x_i \mathsf{P}\lceil x_{m+1} : \phi\rceil)(\bar{b}).\end{aligned}$$

Proof of Lemma 5.8.

Let a be a state in a model $\mathcal{I}$. We need to show $\phi^{\mathcal{I}}(a) = \phi^{\mathcal{I}_a^k}(a)$. Let $\mathcal{J}$ be a new model that extends $\mathcal{I}$ by adding n disjoint copies of both $\mathcal{I}$ and $\mathcal{I}_a^k$. Let $\mathcal{K}$ be the model that extends $\mathcal{I}_a^k$ likewise. We finish the proof by showing that

$$\phi^{\mathcal{I}}(a) = \phi^{\mathcal{J}}(a) = \phi^{\mathcal{K}}(a) = \phi^{\mathcal{I}_a^k}(a).$$

The first and third equality follow by bisimulation invariance of ϕ (Lemma 5.7). The second equality follows by Ehrenfeucht-Fraïssé invariance (Lemma 5.6) once we show that D has a winning strategy in the n-round Ehrenfeucht-Fraïssé game for $\mathcal{J}, a$ and $\mathcal{K}, a$.

Such a winning strategy can be described as follows: For $\bar{a} = (a_1, \dots, a_n)$, put $U^k(\bar{a}) = \bigcup_{i \le n} U^k(a_i)$. Then D maintains the invariant that, if the configuration reached after i rounds is $(\bar{b}, \bar{c})$, then there exists an isomorphism f_i between $U^{k_i}(\bar{b})$ and $U^{k_i}(\bar{c})$ that maps each b_j to the corresponding c_j, where $k_i = 3^{n-i}$.

The invariant holds at the start of the game, because the neighbourhoods on both sides are just $U^k(a)$. Similarly, whenever the invariant holds, the current configuration is a partial isomorphism by restriction of the given isomorphism to the two vectors of the configuration.

Now we consider what happens during the rounds. Suppose that i rounds have been played, and the current configuration is $(\bar{b}, \bar{c})$. If S decides to play a standard round, playing some $b \in \Delta^{\mathcal{J}}$, then there are two cases:

- $b \in U^{2k_{i+1}}(\bar{b})$: In this case, the radius-k_{i+1} neighbourhood $U^{k_{i+1}}(b)$ of b is fully contained in the domain $U^{k_i}(\bar{b})$ of f_i – this follows by the triangle inequality, as $2k_{i+1} + k_{i+1} = 3k_{i+1} = k_i$. Now D can just reply with $c := f_i(b)$, and an isomorphism f_{i+1} between $U^{k_{i+1}}(\bar{b}b)$ and $U^{k_{i+1}}(\bar{c}c)$ is formed by restricting the domain and codomain of f_i appropriately.

- $b \notin U^{2k_{i+1}}(\bar{b})$: In this case, the radius-k_{i+1} neighbourhoods $U^{k_{i+1}}(b)$ of b and $U^{k_{i+1}}(\bar{b})$ of $\bar{b}$ do not intersect – this too follows from the triangle inequality. Now D can pick a fresh copy of $\mathcal{I}$ or $\mathcal{I}_a^k$ in $\mathcal{K}$ (depending on which kind of copy b lies in); her reply c is then just b in that copy. Here, a fresh copy is one that was never visited on any of the previous rounds. By construction of $\mathcal{J}$ and $\mathcal{K}$, such a copy is always available. This means that we now have two isomorphisms, one between $U^{k_{i+1}}(\bar{b})$ and $U^{k_{i+1}}(\bar{c})$ (by restriction of f_i), and one between $U^{k_{i+1}}(b)$ and $U^{k_{i+1}}(c)$ (by isomorphism of the respective copies of $\mathcal{I}$ or $\mathcal{I}_a^k$). Because these isomorphisms have disjoint domains and codomains, we can combine them to form the desired isomorphism f_{i+1}.

If S plays a standard round with some $c \in \Delta^{\mathcal{K}}$ instead, the same argument applies.

Finally, if S starts a probabilistic round by picking an index $0 \le j \le i$ and playing some $\phi_B \colon \Delta^{\mathcal{J}} \to [0,1]$, then we first note that, by the rules of the game, the support of ϕ_B must be contained in $U^1(\bar{b})$, which in turn must be contained in the domain of f_i. This means that D can construct $\phi_C \colon \Delta^{\mathcal{K}} \to [0,1]$ by mapping along f_i, i.e. $\phi_C(c) = \phi_B(f_i^{-1}(c))$ for all successors c of c_j, and $\phi_C(c) = 0$ otherwise. Now, whichever b or c is picked by S, D can just reply with $c := f_i(b)$ or $b := f_i^{-1}(c)$ and f_{i+1} is formed as in the first case of a standard round. Again, the same argument applies if S picks a fuzzy subset ϕ_C on the other side.

Proof of Lemma 6.2.
D wins by maintaining the invariant that the configuration of the game is of the form $(\bar{a}, \mathsf{last}(\bar{a}), 0)$ for some $\bar{a} \in (\Delta^{\mathcal{I}})^+$. To do so, she can put $\mu(\bar{a}a, a) = \pi_{\bar{a}}(\bar{a}a) = \pi_{\mathsf{last}(\bar{a})}(a)$ for all $a \in (\Delta^{\mathcal{I}})^+$, all other values of μ are 0, and $\epsilon' = 0$. Then any move by S leads to a configuration where the invariant holds.

Proof of Lemma 6.3.
Let $\mathcal{I}$ and $\mathcal{J}$ be two models and let $a \in \Delta^{\mathcal{I}}$ and $b \in \Delta^{\mathcal{J}}$ be two states such that $d_k^G(a, b) < \epsilon$. It is enough to show that $|\phi^{\mathcal{I}}(a) - \phi^{\mathcal{J}}(b)| \le \epsilon$.

We denote by a' and a'' the copies of a in $\mathcal{I}^*$ and $(\mathcal{I}^*)_a^k$, respectively. Similarly, b' and b'' denote the copies of b in $\mathcal{J}^*$ and $(\mathcal{J}^*)_b^k$. By Lemma 6.2, D wins the 0-bisimulation-game for $\mathcal{I}, a$ and $\mathcal{I}^*, a'$ (similarly for $\mathcal{J}$) and by Lemma 5.2, she also wins the k-round 0-bisimulation game for $\mathcal{I}^*, a'$ and $(\mathcal{I}^*)_a^k, a''$ (similarly for $\mathcal{J}$). Because behavioural distance d_k^G is a pseudometric, this means that

$$\begin{aligned} &d_k^G(a'', b'') \\ &\le d_k^G(a'', a') + d_k^G(a', a) \\ &\quad + d_k^G(a, b) + d_k^G(b, b') + d_k^G(b', b'') \\ &= d_k^G(a, b) < \epsilon, \end{aligned}$$

so D has a winning strategy in the k-round ϵ-bisimulation game for $(\mathcal{I}^*)_a^k, a''$ and $(\mathcal{J}^*)_b^k, b''$.

In both $(\mathcal{I}^*)_a^k, a''$ and $(\mathcal{J}^*)_b^k, b''$, the reachable states form a tree of depth at most k. This implies that, after i rounds of the game, the two states on either side of the current configuration are nodes at distance i from the root of their respective tree. Thus, whenever k rounds have been played in the game, S does not have a legal move in the next round, because at that point, both nodes in the configuration are necessarily leaves and thus blocking. This in turn means that if D can win the k-round game, she also wins the unbounded game, so, by bisimulation invariance of ϕ, $|\phi^{(\mathcal{I}^*)_a^k}(a'') - \phi^{(\mathcal{J}^*)_b^k}(b'')| \le \epsilon$.

By locality and bisimulation invariance of ϕ, and again Lemma 6.2, we have $\phi^{(\mathcal{I}^*)_a^k}(a'') = \phi^{\mathcal{I}^*}(a') = \phi^{\mathcal{I}}(a)$ as well as $\phi^{(\mathcal{J}^*)_b^k}(b'') = \phi^{\mathcal{J}^*}(b') = \phi^{\mathcal{J}}(b)$. Thus $|\phi^{\mathcal{I}}(a) - \phi^{\mathcal{J}}(b)| \le \epsilon$, as claimed.