On the Expressive Power of First-Order Modal Logic with Two-Dimensional Operators

Alex Kocurek
Group in Logic and the Methodology of Science
University of California, Berkeley
akocurek@berkeley.edu

Abstract. Many authors have noted that there are types of English modal sentences that cannot be formalized in the language of basic first-order modal logic. Some widely discussed examples include "There could have been things other than there actually are" and "Everyone who's actually rich could have been poor." In response to this lack of expressive power, many authors have discussed extensions of first-order modal logic with two-dimensional operators. But claims about the relative expressive power of these extensions are often justified only by example rather than by rigorous proof. In this paper, we provide proofs of many of these claims and present a more complete picture of the expressive landscape for such languages.

---

§1 Introduction

It is well known that first-order modal logic faces fundamental limitations in expressive power. Some standard examples used to illustrate this include:

(E) There could have been things other than there actually are.[1]

(R) Everyone who's actually rich could have been poor.[2]

Informally, the first is true iff there is a possible world where something exists that does not exist in the actual world. The second has multiple readings, but on one reading, it is true iff there is a possible world where everyone who is rich in the actual world is poor in that world. It can been shown that no formula in basic first-order modal logic with actualist quantifiers (i.e., quantifiers ranging over existents) is equivalent to (E) or to (R).[3] We can regiment (E) using a possibilist quantifier Σ (i.e., a quantifier ranging over all possible objects) and an existence predicate E as follows:

$$\Sigma x \left(\neg \mathsf{E}(x) \land \Diamond \mathsf{E}(x) \right). \tag{1}$$

But one can prove that even with these additions, there is still no formula that is equivalent to (R).[4]

---

[1]Originally from Hazen [1976, p. 31].
[2]Originally from Cresswell [1990, p. 34].
[3]Hodes [1984c].
[4]Wehmeier [2001].

In response to these expressive limitations, many authors have considered extending first-order modal logic with an "actually" operator @.[5] They then point out that in the presence of @ and the possibilist universal quantifier Π, the following is equivalent to (R):

$$\Diamond \Pi x \; (@\mathsf{Rich}(x) \to \mathsf{Poor}(x)) \, . \tag{2}$$

However, even with possibilist quantifiers and the actuality operator, sentences like the following seem to remain inexpressible:[6]

(NR) Necessarily, everyone who's rich could have been poor.

One could try to fix this problem by adding more and more operators to the language, some of which we will discuss below. But many such languages face further expressivity limitations themselves.[7] Corresponding expressive limitations also arise for first-order temporal logic, though we will mostly focus on the modal versions until the end of this paper.

Very often, these inexpressibility claims are justified in the literature only by example: all of the most straightforward attempts at formalizing these English sentences into first-order modal logic fail. While this style of argument may be convincing, it does not constitute a proof. One can sometimes find rigorous proofs for a variety of inexpressibility claims.[8] But only Hodes [1984a,b,c] provides proofs of the inexpressibility of (R), (E), and sentences like them in extensions of first-order modal logic with two-dimensional operators such as @.[9] And while these proofs are very interesting and involve a number of underappreciated techniques, they are quite complicated and difficult to generalize to other formal languages of interest.

In this paper, I will use a modular notion of bisimulation to characterize the expressive power of extensions of first-order modal logic with two-dimensional operators. After reviewing basic first-order modal logic (§2), I will provide a single proof method for characterizing the expressive power of a wide variety of first-order modal languages using bisimulations (§3). I will then present a variety of inexpressibility proofs using this technique (§4). I will conclude by generalizing these results to temporal logics and higher-dimensional logics (§5). The more intricate details are left to appendices (§A–D).

§2   First-Order Modal Logic

In this section, we review the standard possible worlds semantics for first-order modal logic. The technical details below are fairly standard, with the exception that our points of evaluation need to be two-dimensional to account for operators like the actuality operator @. While we have picked a particularly simple formulation of first-order modal logic, the

---

[5]Crossley and Humberstone [1977]; Davies and Humberstone [1980]; Hazen [1976, 1990]; Hodes [1984a,b].
[6]Hazen [1976]; Bricker [1989]; Cresswell [1990]; Sider [2010].
[7]van Benthem [1977]; Gabbay [1981]; Cresswell [1990].
[8]See Fine [1979]; Hodes [1984a,b,c]; Forbes [1989]; Wehmeier [2001]; Fritz [2012]; Yanovich [2015].
[9]Forbes [1989, pp. 42–44] gives some inexpressibility results for first-order modal logic with @. See also Correia [2007]; Fritz [2012]; Yanovich [2015] for work on related languages.

inexpressibility results we explore in this paper apply to a wide range of formulations of first-order modal logic.[10]

The signature for our plain vanilla first-order modal language $\mathcal{L}^{1M}$ contains:

- VAR = $\{x_1, x_2, x_3, \ldots\}$ (the set of *(object) variables*);

- $\text{PRED}^n = \{P_1^n, P_2^n, P_3^n, \ldots\}$ for each $n \geqslant 1$ (the set of *n-place predicates*);

The set of *formulas in $\mathcal{L}^{1M}$* or *$\mathcal{L}^{1M}$-formulas* is defined recursively:

$$\varphi ::= P^n(y_1, \ldots, y_n) \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box\varphi \mid \forall x\,\varphi$$

where $P^n \in \text{PRED}^n$ for any $n \geqslant 1$, and $x, y_1, \ldots, y_n \in \text{VAR}$. The usual abbreviations for $\bot$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\exists$, and $\Diamond$ apply. We may drop parentheses for readability. If the free variables of φ are among $y_1, \ldots, y_n$, we may write "$\varphi(y_1, \ldots, y_n)$" to indicate this.

To talk more easily about extensions of $\mathcal{L}^{1M}$, we will introduce a convention. Let $S_1, \ldots, S_n$ be some new symbols with pre-defined syntax. The language obtained from $\mathcal{L}^{1M}$ by adding $S_1, \ldots, S_n$ is $\mathcal{L}^{1M}(S_1, \ldots, S_n)$. Some symbols that might be added include:

$$\varphi ::= \cdots \mid x \approx y \mid \mathsf{E}(y) \mid @\varphi \mid {\downarrow}\varphi \mid \mathcal{F}\varphi \mid \forall_@ x\,\varphi \mid \Pi x\,\varphi$$

where $\approx$ is the identity relation, E is an existence predicate, $@$ is an "actually" operator, $\downarrow$ is a diagonalization operator (the inverse of $@$),[11] $\mathcal{F}$ is a "fixedly" operator,[12] $\forall_@$ is a quantifier over all actual objects,[13] and Π is the possibilist universal quantifier (its existential counterpart is Σ). The usual abbreviations apply. In what follows, we will let "$\mathcal{L}$" stand for any arbitrary $\mathcal{L}^{1M}(S_1, \ldots, S_n)$ where $S_1, \ldots, S_n$ are among the symbols above.

Definition 2.1 *(Models).* A **model** is a tuple $\mathcal{M} = \langle W, R, F, D, \delta, I \rangle$ where:

- W is a nonempty set (the **state space**);

- $R \subseteq W \times W$ (the $\Box$-**accessibility relation**);

- $F \subseteq W \times W$ (the $\mathcal{F}$-**accessibility relation**);

- D is a nonempty set disjoint from W (the **(global) domain**);

- $\delta\colon W \to \wp(D)$ is a function (the **local domain assignment**), where for each $w \in W$, $\delta(w)$ is the **local domain of** w;

- $I\colon \text{PRED}^n \times W \to \wp(D^n)$ (for all $n \geqslant 1$) is a function (the **interpretation function**).

We will let $\mathcal{M}$'s state space be $W^{\mathcal{M}}$, $\mathcal{M}$'s $\Box$-accessibility relation be $R^{\mathcal{M}}$, etc. We will define $R[w] := \{v \in W \mid wRv\}$ (and likewise for F). If $R = W \times W$, we will say R is **universal**. If $D = \bigcup_{w \in W} \delta(w)$ (i.e., D does not contain impossible objects), we will say D satisfies **the domain constraint**. We will let **U** be the class of models where R and F are universal, **D** be the class of models where D satisfies the domain constraint, and **UD** be their intersection.

---

[10]See Garson [2001] for a tree of such formulations.
[11]Vlach [1973]; Lewis [1973].
[12]Davies and Humberstone [1980].
[13]Bricker [1989]; Gilbert [2015].

Let $\mathcal{M}$ be a model. A ***variable assignment for*** $\mathcal{M}$ is a function g mapping variables to elements in D. Let the set of variable assignments for $\mathcal{M}$ be $\mathsf{VA}(\mathcal{M})$. If $g \in \mathsf{VA}(\mathcal{M})$, then g_a^x is the result of modifying g by mapping x to a.

For readability, if $\alpha_1, \ldots, \alpha_n$ is a sequence (of terms, objects, etc.), we may write "$\overline{\alpha}$" in place of "$\alpha_1, \ldots, \alpha_n$". $\overline{\alpha}$ is assumed to be of the appropriate length, whatever that is in a given context. Let $|\overline{\alpha}|$ be the length of $\overline{\alpha}$. When f is some unary function, we may write "$f(\overline{\alpha})$" in place of "$f(\alpha_1), \ldots, f(\alpha_n)$". For instance, if g is a variable assignment, "$g(\overline{x})$" on this notation stands for "$g(x_1), \ldots, g(x_n)$". Likewise, "$g_{\overline{a}}^{\overline{x}}$" stands for "$g_{a_1, \ldots, a_n}^{x_1, \ldots, x_n}$".

Since we want to consider operators like @, our possible worlds semantics will be two-dimensional (as suggested in, e.g., Davies and Humberstone [1980, pp. 4-5]). That is, indices will have to contain two worlds. The first world is to be interpreted as the world "considered as actual" and the second as the world of evaluation.

Definition 2.2 *(Satisfaction for $\mathcal{L}^{1M}$)*. The ***$\mathcal{L}$-satisfaction relation*** $\Vdash$ is defined recursively, for all models $\mathcal{M} = \langle W, R, F, D, \delta, I \rangle$, all $w, v \in W$ and all $g \in \mathsf{VA}(\mathcal{M})$:

$$
\begin{aligned}
\mathcal{M}, w, v, g &\Vdash P^n(\overline{x}) &\Leftrightarrow\quad& \langle g(\overline{x}) \rangle \in I(P^n, v) \\
\mathcal{M}, w, v, g &\Vdash x \approx y &\Leftrightarrow\quad& g(x) = g(y) \\
\mathcal{M}, w, v, g &\Vdash \mathsf{E}(x) &\Leftrightarrow\quad& g(x) \in \delta(v) \\
\mathcal{M}, w, v, g &\Vdash \neg\varphi &\Leftrightarrow\quad& \mathcal{M}, w, v, g \nVdash \varphi \\
\mathcal{M}, w, v, g &\Vdash \varphi \wedge \psi &\Leftrightarrow\quad& \mathcal{M}, w, v, g \Vdash \varphi \text{ and } \mathcal{M}, w, v, g \Vdash \psi \\
\mathcal{M}, w, v, g &\Vdash \Box\varphi &\Leftrightarrow\quad& \forall v' \in R[v]\colon\ \mathcal{M}, w, v', g \Vdash \varphi \\
\mathcal{M}, w, v, g &\Vdash @\varphi &\Leftrightarrow\quad& \mathcal{M}, w, w, g \Vdash \varphi \\
\mathcal{M}, w, v, g &\Vdash {\downarrow}\varphi &\Leftrightarrow\quad& \mathcal{M}, v, v, g \Vdash \varphi \\
\mathcal{M}, w, v, g &\Vdash \mathcal{F}\varphi &\Leftrightarrow\quad& \forall w' \in F[w]\colon\ \mathcal{M}, w', v, g \Vdash \varphi \\
\mathcal{M}, w, v, g &\Vdash \forall x\varphi &\Leftrightarrow\quad& \forall a \in \delta(v)\colon\ \mathcal{M}, w, v, g_a^x \Vdash \varphi \\
\mathcal{M}, w, v, g &\Vdash \forall_@ x\varphi &\Leftrightarrow\quad& \forall a \in \delta(w)\colon\ \mathcal{M}, w, v, g_a^x \Vdash \varphi \\
\mathcal{M}, w, v, g &\Vdash \Pi x\varphi &\Leftrightarrow\quad& \forall a \in D\colon\ \mathcal{M}, w, v, g_a^x \Vdash \varphi.
\end{aligned}
$$

If $|\overline{x}| \leqslant |\overline{a}|$, then $\mathcal{M}, w, v \Vdash \varphi[\overline{a}]$ if for all $g \in \mathsf{VA}(\mathcal{M})$, $\mathcal{M}, w, v, g_{\overline{a}}^{\overline{x}} \Vdash \varphi(\overline{x})$.

Definition 2.3 *(Validity)*. Let $\mathbf{C}$ be a class of models. We will say φ is ***(generally) $\mathbf{C}$-valid***—written as $\Vdash_{\mathbf{C}} \varphi$—if $\mathcal{M}, w, v, g \Vdash \varphi$ for all $\mathcal{M} \in \mathbf{C}$, all $w, v \in W^{\mathcal{M}}$, and all $g \in \mathsf{VA}(\mathcal{M})$. We will say φ is ***diagonally $\mathbf{C}$-valid***—written as $\Vdash_{\mathbf{C}}^{\mathrm{d}} \varphi$—if $\mathcal{M}, w, w, g \Vdash \varphi$ for all $\mathcal{M} \in \mathbf{C}$, all $w \in W^{\mathcal{M}}$, and all $g \in \mathsf{VA}(\mathcal{M})$. If $\varphi \leftrightarrow \psi$ is (diagonally) $\mathbf{C}$-valid, we will say φ and ψ are ***(diagonally) $\mathbf{C}$-equivalent***. If $\mathbf{C}$ is the class of all models, we may drop mention of $\mathbf{C}$ and just say "valid" or "equivalent".

We could have defined some of the additional symbols above in terms of others, assuming the others are present. For instance, $\mathsf{E}(x) \leftrightarrow \exists y\,(x \approx y)$, $\forall x\varphi \leftrightarrow \Pi x\,(\mathsf{E}(x) \to \varphi)$, and $\forall_@ x\varphi \leftrightarrow \Pi x\,(@\mathsf{E}(x) \to \varphi)$ are all valid (we will invoke these throughout without explicit mention of them). Thus, by the following lemma, we could have taken the lefthand

side of these biconditionals to be abbreviations for their righthand side:

Lemma 2.4 *(Replacement of Equivalents).* Suppose φ and ψ have the same free variables. Let θ' be a formula that results from replacing any number of instances of φ in θ with ψ. Then $\Vdash_{\mathbf{C}} \varphi \leftrightarrow \psi$ implies $\Vdash_{\mathbf{C}} \theta \leftrightarrow \theta'$.

This follows by a straightforward induction. If we replace $\Vdash_{\mathbf{C}}$ in **Lemma 2.4** with $\Vdash_{\mathbf{C}}^{\mathbf{d}}$, then the result no longer holds (for instance, $\Vdash^{\mathbf{d}} @P(x) \leftrightarrow P(x)$, but $\nVdash^{\mathbf{d}} \Box @P(x) \leftrightarrow \Box P(x)$). However, if $\Vdash_{\mathbf{C}}^{\mathbf{d}} \varphi \leftrightarrow \psi$, then $\Vdash_{\mathbf{C}} \star\varphi \leftrightarrow \star\psi$, where $\star \in \{@, \downarrow\}$, in which case we can replace $\star\varphi$ with $\star\psi$.

We can think of **Definition 2.2** as specifying a *translation* from the modal language into the language of possible worlds. We can make this more precise by formally defining the language of possible worlds, which is often called the ***correspondence language***.[14] The correspondence language is a two-sorted first-order language: one sort for objects, and one sort for worlds. The signature for our two-sorted first-order language $\mathcal{L}^{\mathsf{TS}}$ contains VAR, PRED, and:

- SVAR $= \{s_1, s_2, s_3, \ldots\}$ (the set of ***state variables***).

The set of ***formulas in $\mathcal{L}^{\mathsf{TS}}$*** or ***$\mathcal{L}^{\mathsf{TS}}$-formulas*** is defined recursively:

$$\alpha ::= P^n(\bar{y}; s) \mid x \approx y \mid s \approx t \mid \mathsf{E}(y; s) \mid \mathsf{R}(s, t) \mid \mathsf{F}(s, t) \mid \neg\alpha \mid (\alpha \wedge \alpha) \mid \forall x\, \alpha \mid \forall s\, \alpha$$

where $P^n \in \mathsf{PRED}^n$, $x, y, \bar{y} \in \mathsf{VAR}$, and $s, t \in \mathsf{SVAR}$. I will typically use $\alpha, \beta, \gamma, \ldots$ for $\mathcal{L}^{\mathsf{TS}}$-formulas to distinguish them from $\mathcal{L}^{\mathsf{1M}}$-formulas.

To illustrate, here are the intended formalizations of (E), (R), and (NR), where s^* is meant to be interpreted as the actual world (which we will assume is the same as our starting world of evaluation, just for simplicity):[15]

(E)   There could have been things other than there actually are.

$$\exists t\, (\mathsf{R}(s^*, t) \wedge \exists x\, (\mathsf{E}(x; t) \wedge \neg\,\mathsf{E}(x; s^*))) \tag{3}$$

(R)   Everyone who's actually rich could have been poor.

$$\exists t\, (\mathsf{R}(s^*, t) \wedge \forall x\, (\mathsf{Rich}(x; s^*) \rightarrow \mathsf{Poor}(x; t))) \tag{4}$$

(NR) Necessarily, everyone who's rich could have been poor.

$$\forall s\, (\mathsf{R}(s^*, s) \rightarrow \exists t\, (\mathsf{R}(s, t) \wedge \forall x\, (\mathsf{Rich}(x; s) \rightarrow \mathsf{Poor}(x; t)))) \tag{5}$$

---

[14]See Blackburn et al. [2001].

[15]There is a general question as to whether the object quantifier in (R) and (NR) is possibilist or actualist. In general, I assume it should be actualist, in which case the object quantifiers in (4) and (5) should technically be E-bounded. But to avoid clutter, we assume in the background that nothing can be rich or poor unless it exists $(\forall s\, \forall x\, ((\mathsf{Rich}(x; s) \vee \mathsf{Poor}(x; s)) \rightarrow \mathsf{E}(x; s)))$, in which case it does not matter which type of quantifier we use in (4) and (5). None of our models will violate this constraint.

To define $\mathcal{L}^{\mathsf{TS}}$'s semantics, we need to modify the definition of a ***variable assignment for*** $\mathcal{M}$ so that not only do variable assignments map variables to elements of D, but they also map state variables to elements of W. Then satisfaction in $\mathcal{L}^{\mathsf{TS}}$ is just the standard notion of satisfaction for two-sorted first-order logic:

Definition 2.5 *(Satisfaction for $\mathcal{L}^{TS}$)*. The $\mathcal{L}^{\boldsymbol{TS}}$***-satisfaction relation*** $\models$ is defined recursively for all models $\mathcal{M}$ and all $g \in \mathsf{VA}(\mathcal{M})$:

$$\mathcal{M}, g \models P^n(\overline{x}; s) \;\Leftrightarrow\; \langle g(\overline{x}) \rangle \in I(P^n, g(s)) \qquad \mathcal{M}, g \models \mathsf{F}(s, t) \;\Leftrightarrow\; g(t) \in F[g(s)]$$

$$\mathcal{M}, g \models x \approx y \quad\;\Leftrightarrow\; g(x) = g(y) \qquad\qquad\qquad \mathcal{M}, g \models \neg\,\alpha \quad\;\Leftrightarrow\; \mathcal{M}, g \not\models \alpha$$

$$\mathcal{M}, g \models s \approx t \quad\;\Leftrightarrow\; g(s) = g(t) \qquad\qquad\qquad \mathcal{M}, g \models \alpha \wedge \beta \;\Leftrightarrow\; \mathcal{M}, g \models \alpha \text{ and } \mathcal{M}, g \models \beta$$

$$\mathcal{M}, g \models \mathsf{E}(x; s) \;\Leftrightarrow\; g(x) \in \delta(g(s)) \qquad\qquad\;\; \mathcal{M}, g \models \forall x\,\alpha \;\;\Leftrightarrow\; \forall a \in D\colon\; \mathcal{M}, g_a^x \models \alpha$$

$$\mathcal{M}, g \models \mathsf{R}(s, t) \;\;\Leftrightarrow\; g(t) \in R[g(s)] \qquad\qquad\;\; \mathcal{M}, g \models \forall s\,\alpha \;\;\Leftrightarrow\; \forall w \in W\colon\; \mathcal{M}, g_w^s \models \alpha.$$

We say α is ***C-valid***—written as $\models_{\mathbf{C}} \alpha$—if $\mathcal{M}, g \models \alpha$ for all $\mathcal{M} \in \mathbf{C}$ and all $g \in \mathsf{VA}(\mathcal{M})$. Equivalence is defined likewise.

We can now make more precise the thought that **Definition 2.2** is specifying a translation:

Definition 2.6 *(Standard Translation)*. Let φ be a $\mathcal{L}$-formula, and let $s, t \in \mathsf{SVAR}$. The ***standard translation of φ with respect to*** $\langle s, t \rangle$, $\mathsf{ST}_{s,t}(\varphi)$, is defined recursively:

$$\mathsf{ST}_{s,t}(P^n(\overline{x})) \;=\; P^n(\overline{x}; t) \qquad\qquad\qquad \mathsf{ST}_{s,t}(\mathcal{F}\varphi) \;=\; \forall s'\; (\mathsf{F}(s, s') \to \mathsf{ST}_{s',t}(\varphi))$$

$$\mathsf{ST}_{s,t}(x \approx y) \;=\; x \approx y \qquad\qquad\qquad\quad\;\; \mathsf{ST}_{s,t}(@\varphi) \;=\; \mathsf{ST}_{s,s}(\varphi)$$

$$\mathsf{ST}_{s,t}(\mathsf{E}(x)) \;=\; \mathsf{E}(x; t) \qquad\qquad\qquad\quad\;\;\; \mathsf{ST}_{s,t}(\downarrow\varphi) \;=\; \mathsf{ST}_{t,t}(\varphi)$$

$$\mathsf{ST}_{s,t}(\neg\,\varphi) \;=\; \neg\,\mathsf{ST}_{s,t}(\varphi) \qquad\qquad\qquad\; \mathsf{ST}_{s,t}(\forall x\,\varphi) \;=\; \forall x\; (\mathsf{E}(x; t) \to \mathsf{ST}_{s,t}(\varphi))$$

$$\mathsf{ST}_{s,t}(\varphi \wedge \psi) \;=\; (\mathsf{ST}_{s,t}(\varphi) \wedge \mathsf{ST}_{s,t}(\psi)) \quad\; \mathsf{ST}_{s,t}(\forall_@ x\,\varphi) \;=\; \forall x\; (\mathsf{E}(x; s) \to \mathsf{ST}_{s,t}(\varphi))$$

$$\mathsf{ST}_{s,t}(\Box\varphi) \;=\; \forall t'\; (\mathsf{R}(t, t') \to \mathsf{ST}_{s,t'}(\varphi)) \quad\;\; \mathsf{ST}_{s,t}(\Pi x\,\varphi) \;=\; \forall x\, \mathsf{ST}_{s,t}(\varphi)$$

where s' and t' are state variables not occurring anywhere in $\mathsf{ST}_{s,t}(\varphi)$. If Φ is a set of $\mathcal{L}$-formulas, then we will let $\mathsf{ST}_{s,t}(\Phi) = \{\mathsf{ST}_{s,t}(\varphi) \mid \varphi \in \Phi\}$.

The following lemma, which can be proved using a simple induction on formulas, states that ST translates every $\mathcal{L}$-formula into an equivalent $\mathcal{L}^{\mathsf{TS}}$-formula:

Lemma 2.7 *(Translation)*. Let $\mathcal{M}$ be a model, $w, v \in W^{\mathcal{M}}$, $g \in \mathsf{VA}(\mathcal{M})$, $s, t \in \mathsf{SVAR}$, and φ an $\mathcal{L}$-formula. Then $\mathcal{M}, w, v, g \Vdash \varphi$ iff $\mathcal{M}, g_{w,v}^{s,t} \models \mathsf{ST}_{s,t}(\varphi)$.

In other words, **Lemma 2.7** tells us that we can think of "$\mathcal{M}, w, v, g \Vdash \varphi$" as a notational variant of "$\mathcal{M}, g_{w,v}^{s,t} \models \mathsf{ST}_{s,t}(\varphi)$". In what follows, we will implicitly identify an extension of $\mathcal{L}^{\mathsf{1M}}$ with its equivalent fragment of the two-sorted language. The question now is to what extent we can find a translation that goes the other way. To help answer this question, we can define a formal notion of expressivity relative to $\mathcal{L}^{\mathsf{TS}}$ as follows:

Definition 2.8 *(Expressivity).* Let **C** be a class of models. We will say a set of $\mathcal{L}^{\mathsf{TS}}$-formulas Γ **C-expresses** a set of $\mathcal{L}^{\mathsf{TS}}$-formulas Δ if Γ is **C**-equivalent to Δ—that is, for all $\mathcal{M} \in \mathbf{C}$ and all $g \in \mathsf{VA}(\mathcal{M})$, we have that $\mathcal{M}, g \Vdash \Gamma$ iff $\mathcal{M}, g \Vdash \Delta$. If either Γ or Δ are singletons, we can drop the set brackets for readability. Where $\mathcal{L}$ is a fragment of $\mathcal{L}^{\mathsf{TS}}$, we will say Γ is **C-expressible** in $\mathcal{L}$ if there is a set of $\mathcal{L}$-formulas that **C**-expresses Γ. Where $\mathcal{L}_1$ and $\mathcal{L}_2$ are fragments of $\mathcal{L}^{\mathsf{TS}}$, we will say $\mathcal{L}_2$ **C-expresses** $\mathcal{L}_1$ or $\mathcal{L}_1$ is **C-included** in $\mathcal{L}_2$—written as $\mathcal{L}_1 \leqslant_{\mathbf{C}} \mathcal{L}_2$—if for any set of $\mathcal{L}_1$-formulas Γ, there is a set of $\mathcal{L}_2$-formulas Δ that **C**-expresses Γ. We will write $\mathcal{L}_1 <_{\mathbf{C}} \mathcal{L}_2$ if $\mathcal{L}_1 \leqslant_{\mathbf{C}} \mathcal{L}_2$ and $\mathcal{L}_2 \nleqslant_{\mathbf{C}} \mathcal{L}_1$, and $\mathcal{L}_1 \equiv_{\mathbf{C}} \mathcal{L}_2$ if $\mathcal{L}_1 \leqslant_{\mathbf{C}} \mathcal{L}_2$ and $\mathcal{L}_2 \leqslant_{\mathbf{C}} \mathcal{L}_1$.

These definitions apply to extensions of $\mathcal{L}^{\mathsf{1M}}$, viewed as fragments of $\mathcal{L}^{\mathsf{TS}}$. Thus, where Γ is a set of $\mathcal{L}^{\mathsf{TS}}$-formulas, and where $\mathcal{L}$ is an extension of $\mathcal{L}^{\mathsf{1M}}$, we will say Γ is **C-expressible** in $\mathcal{L}$ if there is a set of $\mathcal{L}$-formulas Φ such that Γ is **C**-equivalent to $\mathsf{ST}_{s,t}(\Phi)$. Likewise, if $\mathcal{L}_1$ and $\mathcal{L}_2$ are extensions of first-order modal logic, we will write $\mathcal{L}_1 \leqslant_{\mathbf{C}} \mathcal{L}_2$ if for any set of $\mathcal{L}_1$-formulas Φ, there is a set of $\mathcal{L}_2$-formulas Ψ such that $\mathsf{ST}_{s,t}(\Phi)$ is **C**-equivalent to $\mathsf{ST}_{s,t}(\Psi)$. Similarly for $<_{\mathbf{C}}$ and $\equiv_{\mathbf{C}}$.

§3   BISIMULATION

To show that no formula (or set of formulas) of a modal language $\mathcal{L}$ can express a certain formula α of $\mathcal{L}^{\mathsf{TS}}$, one must generally construct two models such that (a) they agree in $\mathcal{L}$ on all $\mathcal{L}$-formulas (i.e., they are $\mathcal{L}$-*equivalent*), and (b) they disagree in $\mathcal{L}^{\mathsf{TS}}$ on α. To make showing that such models are $\mathcal{L}$-equivalent easier, we can appeal to the notion of a *bisimulation*.[16] The notion of a bisimulation for first-order modal logic has not been discussed much until recently.[17] Below, we extend the notion of bisimulation in order to ensure modal equivalence for formulas involving two-dimensional operators.

A bisimulation is basically a back-and-forth game. In the standard back-and-forth game for (non-modal) first-order logic, there are two players, Abelard and Eloïse. Abelard aims to refute Eloïse's attempt to show that the two models satisfy the same closed formulas. Abelard starts by picking an object from one of the models. Eloïse must then pick a matching object from the other model that satisfies the same atomic formulas. They continue in this manner, making sure at all times that the objects picked out so far from one model satisfy exactly the same atomic formulas that the objects picked out from the other model satisfy. If at any point the objects picked out from one model do not satisfy the same atomic formulas as the objects picked out from the other model, then Abelard wins. But if Eloïse manages to extend the game out for an infinite number of rounds, she wins. Two first-order models are elementarily equivalent (i.e., satisfy the same closed first-order formulas) if (but not only if) Eloïse has a winning strategy in this game for those models.

Likewise, two modal models satisfy the same $\mathcal{L}^{\mathsf{1M}}$-formulas if Eloïse has a winning strategy for a back-and-forth game like the one above, with some modifications. In the modified game, the game is "located" at some world(s) in the two models. When Abelard

---

[16]See Blackburn et al. [2001, Chp. 2] for an introduction to bisimulations.

[17]See Fine [1981]; Sturm and Wolter [2001]; van Benthem [2010]; Fritz [2012]; Yanovich [2015].

picks an object from the model, he must pick an object that exists at the world where the game is located; likewise with Eloïse. Now the catch: Abelard can choose, at any time, to change the location of the game in either model to any accessible world from the current location. In order to keep playing, Eloïse must likewise pick a matching accessible world in the other model to relocate the game to. The game then relocates to those accessible worlds, and the game continues. As before, if the objects that have been picked out from one model do not satisfy the same atomic formulas *at the game's current location* that are satisfied by the objects picked out from the other model, then Abelard wins. But if Eloïse manages to extend the game out for an infinite number of rounds, she wins. Two worlds in two models will satisfy the same $\mathcal{L}^{1M}$-formulas if Eloïse has a winning strategy in this game, where the game starts at those two worlds. More variations arise when different extensions of $\mathcal{L}^{1M}$ are considered. More precisely:

Definition 3.1 (*Bisimulation*). Let M and N be models. An $\mathcal{L}^{1M}$**-bisimulation between** M **and** N is a nonempty variably polyadic relation Z such that for all $w, v \in W^M$, all $w', v' \in W^N$, all finite $\overline{a} \in D^M$, and all finite $\overline{b} \in D^N$ where $|\overline{a}| = |\overline{b}|$, we have that $Z(w, v, \overline{a}; w', v', \overline{b})$ only if:

(Atomic) $\forall m \in \mathbb{N} \, \forall P^m \in \mathsf{PRED}^m \, \forall i_1, \ldots, i_m \leqslant |\overline{a}|$:

$$\langle a_{i_1}, \ldots, a_{i_m} \rangle \in I^M(P^m, v) \quad \Leftrightarrow \quad \langle b_{i_1}, \ldots, b_{i_m} \rangle \in I^N(P^m, v')$$

(Zig) $\forall u \in R^M[v] \, \exists u' \in R^N[v'] \colon Z(w, u, \overline{a}; w', u', \overline{b})$

(Zag) $\forall u' \in R^N[v'] \, \exists u \in R^M[v] \colon Z(w, u, \overline{a}; w', u', \overline{b})$

(Forth) $\forall a' \in \delta^M(v) \, \exists b' \in \delta^N(v') \colon Z(w, v, \overline{a}, a'; w', v', \overline{b}, b')$

(Back) $\forall b' \in \delta^N(v') \, \exists a' \in \delta^M(v) \colon Z(w, v, \overline{a}, a'; w', v', \overline{b}, b')$.

We may write "$M, w, v, \overline{a} \leftrightarroweq N, w', v', \overline{b}$" (where possibly $|\overline{a}| = |\overline{b}| = 0$) to indicate that $M, w, v, \overline{a}$ and $N, w', v', \overline{b}$ are $\mathcal{L}^{1M}$**-bisimilar**, i.e., that there is a bisimulation Z between M and N such that $Z(w, v, \overline{a}; w', v', \overline{b})$.

   The notion of an $\mathcal{L}^{1M}(S_1, \ldots, S_n)$**-bisimulation between** M **and** N is defined similarly, except one must add the corresponding condition(s) below:

(Eq) $\forall n, m \leqslant |\overline{a}| : a_n = a_m$ iff $b_n = b_m$

(Ex) $\forall n \leqslant |\overline{a}| : a_n \in \delta^M(v)$ iff $b_n \in \delta^N(v')$

(Act) $Z(w, w, \overline{a}; w', w', \overline{b})$

(Diag) $Z(v, v, \overline{a}; v', v', \overline{b})$

($\mathcal{F}$-Zig) $\forall u \in F^M[w] \, \exists u' \in F^N[w'] \colon Z(u, v, \overline{a}; u', v', \overline{b})$

($\mathcal{F}$-Zag) $\forall u' \in F^N[w'] \, \exists u \in F^M[w] \colon Z(u, v, \overline{a}; u', v', \overline{b})$

($\forall_@$-Forth) $\forall a' \in \delta^M(w) \, \exists b' \in \delta^N(w') \colon Z(w, v, \overline{a}, a'; w', v', \overline{b}, b')$

($\forall_@$-Back) $\forall b' \in \delta^N(w) \exists a' \in \delta^M(w') :\ Z(w, v, \overline{a}, a'; w', v', \overline{b}, b')$

(Π-Forth) $\forall a' \in D^M \exists b' \in D^N :\ Z(w, v, \overline{a}, a'; w', v', \overline{b}, b')$

(Π-Back) $\forall b' \in D^N \exists a' \in D^M :\ Z(w, v, \overline{a}, a'; w', v', \overline{b}, b').$

We may write "$M, w, v, \overline{a} \leftrightarrows_{\mathcal{L}} N, w', v', \overline{b}$" to indicate that $M, w, v, \overline{a}$ and $N, w', v', \overline{b}$ are $\mathcal{L}$-bisimilar. We may also sometimes write "$M, w, v, \overline{a} \leftrightarrows_{S_1,\ldots,S_n} N, w', v', \overline{b}$", where $\mathcal{L} = \mathcal{L}^{1M}(S_1, \ldots, S_n)$, for readability.

Here are the various conditions phrased in terms of games. (Atomic) says that Eloïse loses unless $\overline{a}$ satisfy the same atomic formulas in M, w, v that $\overline{b}$ satisfy in N, w', v'. (Zig) says that if Abelard decides to move the game to $\langle w, u \rangle$ in M where $u \in R^M[v]$, Eloïse must choose a $u' \in R^N[v']$ and relocate the game in N to $\langle w', u' \rangle$ to continue. Likewise for (Zag). (Forth) says that if Abelard picks an object a' from v, Eloïse must pick an object b' from v' to match it with. Likewise for (Back). (Eq) says that if Abelard picks an object that was already chosen, Eloïse must pick the matching object. (Ex) says that the objects picked have to agree in terms of existence, even when the game relocates. (Act) says that Abelard can force the game to relocate to $\langle w, w \rangle$ and $\langle w', w' \rangle$. Likewise for (Diag). The other clauses are as before, except with respect to different domains and relations.

Definition 3.2 (*Modal Equivalence*). We will say that $M, w, v, \overline{a}$ and $N, w', v', \overline{b}$ are $\mathcal{L}$-*equivalent* or **modally equivalent** if for all $\mathcal{L}$-formulas $\varphi(\overline{x})$ (where $|\overline{x}| \leqslant |\overline{a}|$), $M, w, v \Vdash \varphi[\overline{a}]$ iff $N, w', v' \Vdash \varphi[\overline{b}]$. We may write "$M, w, v, \overline{a} \equiv_{\mathcal{L}} N, w', v', \overline{b}$" to indicate that $M, w, v, \overline{a}$ and $N, w', v', \overline{b}$ are $\mathcal{L}$-equivalent.

Theorem 3.3 (*Bisimulation Implies Modal Equivalence*). Where $\mathcal{L} = \mathcal{L}^{1M}(S_1, \ldots, S_n)$, if $M, w, v, \overline{a} \leftrightarrows_{\mathcal{L}} N, w', v', \overline{b}$, then $M, w, v, \overline{a} \equiv_{\mathcal{L}} N, w', v', \overline{b}$.

In general, modal equivalence does not imply bisimulation.[18] However, it does when infinitary conjunction is present in the language. Consider the symbol $\bigwedge$ with the following formation rule: if Φ is a set of well-formed formulas (of any size), then $\bigwedge \Phi$ is a well-formed formula. Then $M, w, v, a \equiv_{\bigwedge, S_1,\ldots,S_n} N, w', v', \overline{b}$ iff $M, w, v, \overline{a} \leftrightarrows_{S_1,\ldots,S_n} N, w', v', \overline{b}$.[19] Thus, bisimulation is equivalent to infinitary modal equivalence. No bisimulation clauses need to be added for $\bigwedge$.

Adding infinitary conjunction to the language clearly increases the expressive power of the language. For example, one can regiment the sentence "There are infinitely many rich people" as $\bigwedge_{n \in \omega} \exists_{\geqslant n} x \operatorname{Rich}(x)$, where $\bigwedge_{n \in \omega} \varphi_n$ is short for the formula $\bigwedge \{ \varphi_n \mid n \in \omega \}$, and $\exists_{\geqslant n} x \varphi(x)$ is short for the formula $\exists x_1 \cdots \exists x_n \left(\bigwedge_{i=1}^{n} \varphi(x_i) \wedge \bigwedge_{i \neq j} x_i \not\approx x_j \right)$. However, infinitary conjunction does not increase the expressive power enough to overcome the particular expressive limitations discussed here, so we set it aside in what follows.

---

[18]See Blackburn et al. [2001, p. 68] for the proof in the propositional case.

[19]See Goranko and Otto [2007] for a proof in the propositional case. Generalizing to first-order modal logic is straightforward.

Now, recall the definition of expressibility (**Definition 2.8**).

Corollary 3.4 *(Translation Implies Invariance)*. Let $\alpha(\overline{x}; s, t)$ be an $\mathcal{L}^{\text{TS}}$-formula. Given $\mathcal{M}, w, v, \overline{a} \leftrightarrows_{S_1,\dots,S_n} \mathcal{N}, w', v', \overline{b}$, and given α is equivalent to the translation of some set of $\mathcal{L}^{1\text{M}}(S_1, \dots, S_n)$-formulas, then $\mathcal{M} \vDash \alpha[\overline{a}; w, v]$ iff $\mathcal{N} \vDash \alpha[\overline{b}; w', v']$. In other words, if $\mathcal{M}, w, v, \overline{a}$ and $\mathcal{N}, w', v', \overline{b}$ are $\mathcal{L}^{1\text{M}}(S_1, \dots, S_n)$-bisimilar, but they disagree on α, then α is not expressible as a set of $\mathcal{L}^{1\text{M}}(S_1, \dots, S_n)$-formulas.

Corollary 3.4 says that if a $\mathcal{L}^{\text{TS}}$-formula is equivalent to the translation of an $\mathcal{L}$-formula (or a set of $\mathcal{L}$-formulas), then it is preserved under $\mathcal{L}$-bisimulations. As in propositional modal logic, the converse also holds (see §A for the proof).[20]

Theorem 3.5 *(van Benthem Characterization Theorem)*. Let $\alpha(\overline{x}; s, t)$ be an $\mathcal{L}^{\text{TS}}$-formula such that $\mathcal{M} \vDash \alpha[\overline{a}; w, v]$ iff $\mathcal{N} \vDash \alpha[\overline{b}; w', v']$ given that $\mathcal{M}, w, v, \overline{a} \leftrightarrows_{\mathcal{L}} \mathcal{N}, w', v', \overline{b}$. Then α is equivalent to the translation of some $\mathcal{L}$-formula.

This together with **Theorem 3.3** implies that $\mathcal{L}$ is just the $\mathcal{L}$-bisimulation invariant fragment of $\mathcal{L}^{\text{TS}}$. For our purposes, however, **Corollary 3.4** will be the key result in generating the inexpressibility results below.

§4 INEXPRESSIBILITY

While **Corollary 3.4** and **Theorem 3.5** exactly characterize the expressive power of $\mathcal{L}^{1\text{M}}$ and its various extensions, the characterization is a bit abstract, and it does not automatically tell us what the expressive power of these extensions are relative to one another. We now turn to illustrating the expressive limitations of $\mathcal{L}^{1\text{M}}$ and its extensions with concrete examples. Note that all of our models in this section fall in the class **UD**, so these inexpressibility results therefore apply to any class that includes **UD**.

To warm up, we start by showing that (E) is not expressible in $\mathcal{L}^{1\text{M}}$. Recall (E) says that there could have been things other than there are, which is formalized in $\mathcal{L}^{\text{TS}}$ as (3):

$$\exists t \, (\mathsf{R}(s^*, t) \wedge \exists x \, (\mathsf{E}(x; t) \wedge \neg \, \mathsf{E}(x; s^*))). \tag{3}$$

The proof strategy will always be the same: construct two modal models that are bisimilar, but that disagree on (3). Because we do not have $\approx$ in $\mathcal{L}^{1\text{M}}$ by default, this is actually very easy. Let $\mathcal{E} = \langle W, R, F, D, \delta, I \rangle$, where $W = \{w, v\}$, R and F are universal, $D = \delta(w) = \delta(v) = \{a\}$, and $I(P, u) = \varnothing$. Let $\mathcal{E}' = \langle W, R, F, D', \delta', I \rangle$, where everything is as in $\mathcal{E}$, except $D' = \delta'(v) = \{a, b\}$. See Figure 1 for a picture.

It is easy to see that w in $\mathcal{E}$ does not satisfy (3): every possible object exists at w. However, w in $\mathcal{E}'$ does satisfy (3): b is a possible object that does not exist at w. So $\mathcal{E}, w$ and $\mathcal{E}', w$ disagree on (3). So we just need to show that $\mathcal{E}, w, w \leftrightarrows \mathcal{E}', w, w$. In fact, in this case, we can just take Z to map every world to every world, and every object to every object any

---

[20]The proof is essentially the same as the proof for propositional modal logic. See Blackburn et al. [2001, Chp. 2.6] and Sturm and Wolter [2001, pp. 579–580].
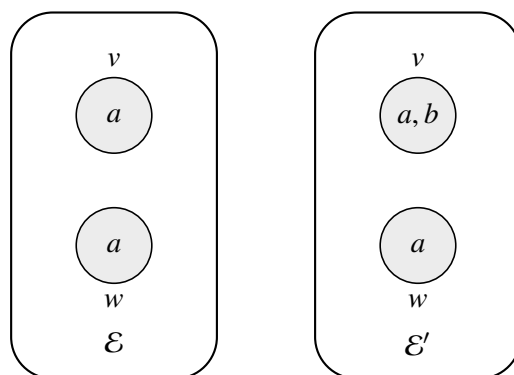
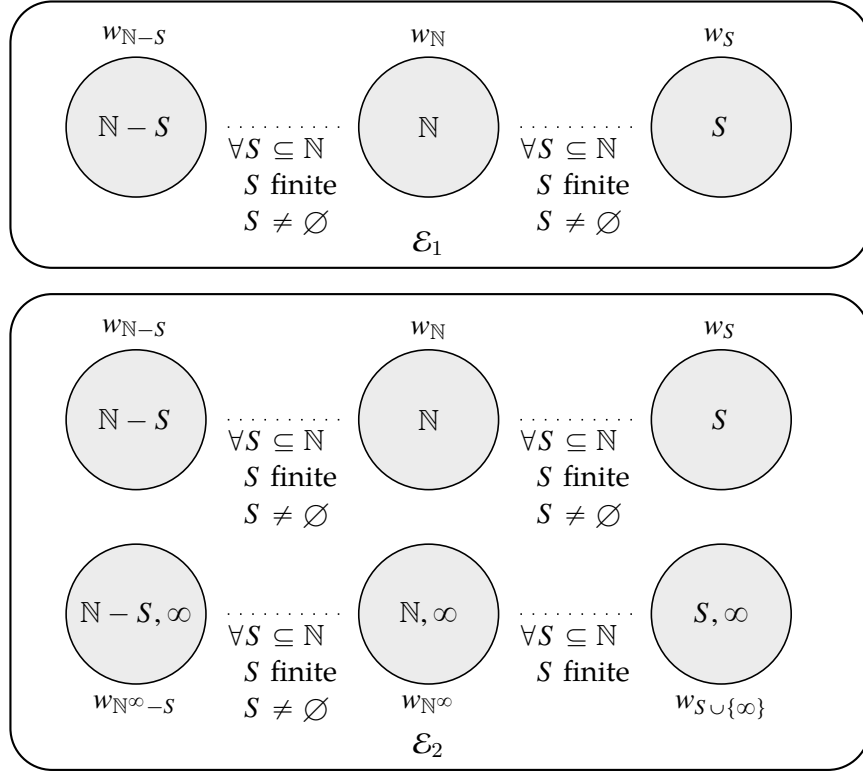Figure 1: $\mathcal{L}^{1M}$-bisimilar models that disagree on (E).

number of times. To show this is a bisimulation, we just check each of the clauses from **Definition 3.1** holds, which is easy to do. One might initially think that we will run into problems in trying to show (Back) holds; for if we consider $Z(w, v, a; w, v, a)$, and we decide to pick b from $\delta'(v)$, then we cannot pick b from $\delta(v)$ to match it with. But since a and b do not disagree on any predicates, and since $\approx$ is not present in the language, $\mathcal{L}^{1M}$ cannot tell that a and b are distinct objects anyway. We do not have to match a to a and b to b every time. We can just as well match b in $\delta'(v)$ with a in $\delta(v)$.

What made this proof easy was the absence of $\approx$. Now we will show that even $\mathcal{L}^{1M}(\approx)$ cannot express (E).[21] Consider first $\mathcal{E}_1 = \langle W_1, W_1^2, W_1^2, \mathbb{N}, \delta_1, I_1 \rangle$, where the global domain is $\mathbb{N}$ and the accessibility relations are both universal. For each nonempty finite or cofinite $S \subseteq \mathbb{N}$, there is a world $w_S \in W_1$ such that $\delta_1(w_S) = S$. No other worlds are in W_1. Again, the extension of all non-logical predicates will be empty at all worlds. The second model $\mathcal{E}_2 = \langle W_2, W_2^2, W_2^2, \mathbb{N} \cup \{\infty\}, \delta_2, I_2 \rangle$ is similar to the first, but now the global domain contains an additional object ∞. For each nonempty set S that is either finite or cofinite in $\mathbb{N}^\infty :=$ $\mathbb{N} \cup \{\infty\}$, there is a world $w_S \in W_2$ such that $\delta_2(w_S) = S$. See Figure 2 for a picture.

---

[21] A proof of this was suggested by Hazen [1976, p. 35]. He describes his models as follows:

> For suppose that [(3)] is false, that the actual world is the only one with infinitely many individuals, and that for every finite set of individuals in the actual world there is a world containing just those individuals, and consider the purely logical sentences true under those suppositions. Now suppose there is added to the system of possible worlds a new world for each old world, containing all the same individuals plus one new individual (the same for each new world) not in any old world. [(3)] will have become true, but no purely logical sentence of the modal language will have changed its truth value.

However, the proof is not correct as stated, since the second model, but not the first, satisfies the formula $\diamond \exists x \, \square (\mathsf{E}(x) \to \exists z \, (x \not\approx z))$ (there is no world where this new object exists by itself). The natural fix is to add another world to the second model which only contains the new object. The resulting models (which are like ours except the worlds with cofinite domains are removed) satisfy the same $\mathcal{L}^{1M}(\approx)$-formulas, but they are not bisimilar, since they are distinguished by the $\mathcal{L}^{1M}(\bigwedge, \approx)$-formula $\diamond \exists x \diamond (\bigwedge_{n \in \omega} \exists_{\geq n} y \, \mathsf{E}(y) \wedge \neg \mathsf{E}(x))$. However, if you restrict the $\mathcal{L}^{1M}(\approx)$-bisimulation game to n-steps, for any finite n, then Eloïse has a winning strategy (that is, the models are n-bisimilar for every n); and this suffices to guarantee modal equivalence. Here, we ensure full bisimilarity by including worlds with cofinite domains. Thus, our proof shows that (3) is not even expressible in $\mathcal{L}^{1M}(\bigwedge, \approx)$.

Figure 2: $\mathcal{L}^{1\mathsf{M}}(\approx)$-bisimilar models disagreeing on (E).

Since $\delta_1(w_{\mathbb{N}}) = \mathbb{N} = D$, $w_{\mathbb{N}}$ in $\mathcal{E}_1$ does not satisfy (3). But $w_{\mathbb{N}}$ in $\mathcal{E}_2$ does satisfy (3), since $\infty \notin \delta_2(w_{\mathbb{N}})$. So $\mathcal{E}_1, w_{\mathbb{N}}$ and $\mathcal{E}_2, w_{\mathbb{N}}$ disagree on (3). So we just need to show that $\mathcal{E}_1, w_{\mathbb{N}}, w_{\mathbb{N}} \leftrightarroweq_{\approx} \mathcal{E}_2, w_{\mathbb{N}}, w_{\mathbb{N}}$. Constructing the bisimulation is fairly straightforward (albeit tedious) once we work out what Eloïse's winning strategy is. The construction of the bisimulation is given in §B, but the idea in terms of games is sketched below. To help describe the proof, let us introduce the following useful definition:

Definition 4.1 *(Partial Isomorphism).* A ***partial isomorphism between*** $\mathcal{M}, w, v, \bar{a}$ ***and*** $\mathcal{N}, w', v', \bar{b}$ is a finite partial map $\rho \colon D \to D'$ such that $\rho(a_i) = b_i$ for $i \leqslant |\bar{a}|$ and:

(Predicate) $\forall m \in \mathbb{N}\, \forall P^m \in \mathsf{PRED}^m\, \forall c_1, \ldots, c_m \in \mathrm{dom}\,(\rho)$:

$$\langle c_1, \ldots, c_m \rangle \in I^{\mathcal{M}}(P^m, v) \quad \Leftrightarrow \quad \langle \rho(c_1), \ldots, \rho(c_m) \rangle \in I^{\mathcal{N}}(P^m, v').$$

(Existence) $\forall a \in \mathrm{dom}\,(\rho) \colon a \in \delta^{\mathcal{M}}(v)$ iff $\rho(a) \in \delta^{\mathcal{N}}(v')$.

(Equality) $\forall c, d \in \mathrm{dom}\,(\rho) \colon \rho(c) = \rho(d) \Rightarrow c = d$.

We write $\mathcal{M}, w, v, \bar{a} \simeq \mathcal{N}, w', v', \bar{b}$ to indicate that there is a partial isomorphism between $\mathcal{M}, w, v, \bar{a}$ and $\mathcal{N}, w', v', \bar{b}$.

To say that $\mathcal{M}, w, v, \overline{a} \simeq \mathcal{N}, w', v', \overline{b}$ is essentially to say that Eloïse can continue the game (i.e., Abelard has not won yet) at this stage of the game (even with $\approx$ present).

Proposition 4.2 *(Inexpressibility of (E))*. $\mathcal{E}_1, w_\mathbb{N}, w_\mathbb{N} \leftrightarrows_\approx \mathcal{E}_2, w_\mathbb{N}, w_\mathbb{N}$. But $\mathcal{E}_2 \vDash (3)[w_\mathbb{N}]$ while $\mathcal{E}_1 \nvDash (3)[w_\mathbb{N}]$. Hence, (3) is not expressible in $\mathcal{L}^{1M}(\approx)$.

Proof *(Sketch)*:   Our game starts at $\mathcal{E}_1, w_\mathbb{N}, w_\mathbb{N}$ and $\mathcal{E}_2, w_\mathbb{N}, w_\mathbb{N}$. We will describe a strategy for Eloïse such that, at every stage of the game, which we will represent as $\langle w_\mathbb{N}, v_1, \overline{a}; w_\mathbb{N}, v_2, \overline{b} \rangle$, we have that $\mathcal{E}_1, w_\mathbb{N}, v_1, \overline{a} \simeq \mathcal{E}_2, w_\mathbb{N}, v_2, \overline{b}$ (in other words: Eloïse can continue the game at every stage of the game). We construct the strategy by induction on Abelard's move.

Vacuously, $\mathcal{E}_1, w_\mathbb{N}, w_\mathbb{N} \simeq \mathcal{E}_2, w_\mathbb{N}, w_\mathbb{N}$. So suppose $\langle w_\mathbb{N}, v_1, \overline{a}'; w_\mathbb{N}, v_2, \overline{b} \rangle$ is the current stage of the game, where $\mathcal{E}_1, w_\mathbb{N}, v_1, \overline{a} \simeq \mathcal{E}_2, w_\mathbb{N}, v_2, \overline{b}$ and where $|\delta_1(v_1)| = |\delta_2(v_2)|$, i.e., the size of $\delta_1(v_1)$ and $\delta_2(v_2)$ is the same. Abelard can decide either to pick an object from $\delta_1(v_1)$ or $\delta_2(v_2)$, or to relocate the game. By the fact that $\mathcal{E}_1, w_\mathbb{N}, v_1, \overline{a} \simeq \mathcal{E}_2, w_\mathbb{N}, v_2, \overline{b}$ and that $|\delta_1(v_1)| = |\delta_2(v_2)|$, it follows that $|\delta_1(v_1) - \{\overline{a}\}| = |\delta_2(v_2) - \{\overline{b}\}|$. So if Abelard decides to pick a new object from one of v_1 and v_2, Eloïse can always pick a matching object from the local domain of the other world to continue the game.

Suppose instead that Abelard decides to relocate the game. Eloïse should then chose a world in the other model so that the following holds of the new locations u_1 and u_2: (i) $|\delta_1(u_1)| = |\delta_2(u_2)|$, and (ii) $a_i \in \delta_1(u_1)$ iff $b_i \in \delta_2(u_2)$. Since there are only finitely many $\overline{a}$ at any given stage of the game, this will always be possible. And as long as (ii) holds, we will have that $\mathcal{M}, w_\mathbb{N}, u_1, \overline{a} \simeq \mathcal{M}, w_\mathbb{N}, u_2, \overline{b}$.   ∎

Notice that this proof will fail under a variety of extensions of $\mathcal{L}^{1M}(\approx)$. This is easy to see if the extension can express (3) directly, as does $\mathcal{L}^{1M}(\approx, @)$ and $\mathcal{L}^{1M}(\approx, \Pi)$:

$$\Diamond \exists x \, @ \, \neg \, \mathsf{E}(x) \tag{6}$$

$$\Sigma x \, (\Diamond \mathsf{E}(x) \wedge \neg \, \mathsf{E}(x)). \tag{7}$$

But it is also instructive to see where the proof for **Proposition 4.2** fails for these extensions. If we were to add Π, then Abelard would be allowed to pick any object from the global domain of either model. In that case, partial isomorphism is no longer enough to guarantee that Eloïse can continue in this game. For Eloïse to continue the game at $\langle w_\mathbb{N}, v_1, \overline{a}; w_\mathbb{N}, v_2, \overline{b} \rangle$, we also need to ensure that $|\mathbb{N} - \delta_1(v_1)| = |\mathbb{N}^\infty - \delta_2(v_2)|$. Otherwise, if say $|\mathbb{N} - \delta_1(v_1)| < |\mathbb{N}^\infty - \delta_2(v_2)|$, Abelard could keep picking objects from $\mathbb{N} - \delta_1(v_1)$ until he ran out (since $|\mathbb{N} - \delta_1(v_1)| < |\mathbb{N}^\infty - \delta_2(v_2)|$ would imply that $\mathbb{N} - \delta_1(v_1)$ is finite). Then he could pick whatever unmatched objects remain in $\mathbb{N}^\infty - \delta_2(v_2)$. In response, Eloïse would be forced to match Abelard's object either with an object not in $\mathbb{N} - \delta_1(v_1)$, thus violating (Ex) from **Definition 3.1**, or with an object in $\mathbb{N} - \delta_1(v_1)$ that was already chosen, thus matching a previously unmatched object to a previously matched object and violating (Eq). So we need to be able to ensure that $|\mathbb{N} - \delta_1(v_1)| = |\mathbb{N}^\infty - \delta_2(v_2)|$ at every stage of the game, which we can do *except* at one very crucial point, viz., the beginning:

$|\mathbb{N} - \delta_1(w_{\mathbb{N}})| \neq |\mathbb{N}^\infty - \delta_2(w_{\mathbb{N}})|$. In other words, Abelard can force a win just by picking ∞ from D_2, leaving Eloïse unable to pick a matching object while meeting (Ex). Without Π, this winning strategy for Abelard is blocked.

If we add @, then Abelard can force the location of the game in both models to move back to $w_{\mathbb{N}}$. In the proof of **Proposition 4.2** above, it is crucial that Eloïse can choose to relocate to a world similar enough to the actual world. For instance, suppose on round 1, Abelard chooses to move to $w_{\{\infty\}}$ in $\mathcal{E}_2$. Then Eloïse must choose to move to some $w_{\{n\}}$ in $\mathcal{E}_1$—let us say $w_{\{5\}}$. Abelard can then choose ∞ from $\delta_2(w_{\{\infty\}})$, forcing Eloïse to choose 5. At this point, if Abelard decided to relocate the game back to $w_{\mathbb{N}}$ in both models, then he could choose 5 from $\mathcal{E}_1$, and Eloïse would lose by violating (Eq). But without @, while Abelard can choose to relocate the game to $w_{\mathbb{N}}$ in $\mathcal{E}_2$, say, Eloïse does not have to relocate the game to $w_{\mathbb{N}}$ in $\mathcal{E}_1$; she can, for instance, pick to relocate to $w_{\mathbb{N} - \{5\}}$ in $\mathcal{E}_1$.

Let us now turn to showing a more difficult inexpressibility result, viz., that (R) is not expressible in $\mathcal{L}^{1M}(\approx, @)$.[22] Recall (R) says that everyone who is actually rich could have been poor, which is formalized in $\mathcal{L}^{TS}$ as (4):

$$\exists t \ (R(s^*, t) \wedge \forall x \ (\mathsf{Rich}(x; s^*) \rightarrow \mathsf{Poor}(x; t))). \tag{4}$$

Let $\mathbb{N}^- := \mathbb{Z} - \mathbb{N}$. We let $\mathcal{R}_1 = \langle W_1, R_1, F_1, D_1, \delta_1, I_1 \rangle$ and $\mathcal{R}_2 = \langle W_2, R_2, F_2, D_2, \delta_2, I_2 \rangle$, where $D_1 = D_2 = \mathbb{Z}$ and the accessibility relations are universal for both models. There is a world $w \in W_1$ that will act as our actual world, where every positive integer is rich (top half of circle), and every negative integer is poor (bottom half of circle). For each nonempty finite subset $S \subseteq \mathbb{N}$, there is a world $v_S \in W_1$ where the members of S do not exist, and otherwise the rich and the poor are flipped with respect to w; so at v_S, the negative integers are rich, and the positive integers not in S are poor, and the positive integers in S do not exist. The extension of all other predicates is empty. The only difference between $\mathcal{R}_1$ and $\mathcal{R}_2$ is that $\mathcal{R}_2$ includes an additional world $v_\varnothing \in W_2$, where no integer fails to exist, and where the rich and poor are flipped with respect to w. See Figure 3 for a picture.

$\mathcal{R}_2 \vDash (4)[w]$, since $v_\varnothing$ is the world where everyone rich in w is poor. But $\mathcal{R}_1 \nvDash (4)[w]$, since in every world where something that is rich in w is poor, something that is rich in w does not exist (and hence is not poor). And once again, $\mathcal{R}_1, w, w \leftrightarroweq_{\approx, @} \mathcal{R}_2, w, w$. The details are left to §B, but a proof is sketched in terms of games below.

Proposition 4.3 (*Inexpressibility of (R)*). $\mathcal{R}_1, w, w \leftrightarroweq_{\approx, @} \mathcal{R}_2, w, w$. But $\mathcal{R}_2 \vDash (4)[w]$ even though $\mathcal{R}_1 \nvDash (4)[w]$. Hence, (4) is not expressible in $\mathcal{L}^{1M}(\approx, @)$.

---

[22]Yanovich [2015, p. 87] claims to have shown that $\Diamond \Pi x \ (@Q(x) \rightarrow Q(x))$ is not expressible in $\mathcal{L}^{1M}(\approx, \Pi)$. He proceeds, as we do, by constructing two models that disagree on this sentence, and then argues that they are bisimilar. His first model consists of two worlds w and u, where $R = \{\langle w, u \rangle\}$ and where in both w and u, $a_1, a_2, a_3, \ldots$ satisfy $Q(x)$ and $b_1, b_2, b_3, \ldots$ do not satisfy $Q(x)$. His second model consists of worlds w', u'_1, and u'_2, where $R' = \{\langle w', u'_1 \rangle, \langle w', u'_2 \rangle\}$. In w', $c_1, c_2, c_3, \ldots, d_1, d_2, d_3, \ldots$ satisfy $Q(x)$ while $e_1, e_2, e_3, \ldots$ do not. In u'_1, only $c_1, c_2, c_3, \ldots$ satisfy Q, and in u'_2, only $d_1, d_2, d_3, \ldots$ satisfy Q. He then claims that w and w' are $\mathcal{L}^{1M}(\approx, \Pi)$-bisimilar. However, these models are not $\mathcal{L}^{1M}(\approx, \Pi)$-bisimilar. In fact, they do not even satisfy the same $\mathcal{L}^{1M}(\Pi)$-formulas: e.g., $\Sigma x \ (Q(x) \wedge \Diamond \neg Q(x))$ distinguishes the two models. The claim that $\mathcal{L}^{1M}(@, \Pi)$ is not included in $\mathcal{L}^{1M}(\approx, \Pi)$ is still correct, as can be verified with a bisimulation argument using the models $\mathcal{N}_1, w, w$ and $\mathcal{N}_2, w, w$ defined in Figure 4 below (where $w = w^\varnothing_\varnothing$). Another proof that $\mathcal{L}^{1M}(\approx, \Pi)$ cannot express (R) can be found in Wehmeier [2001].
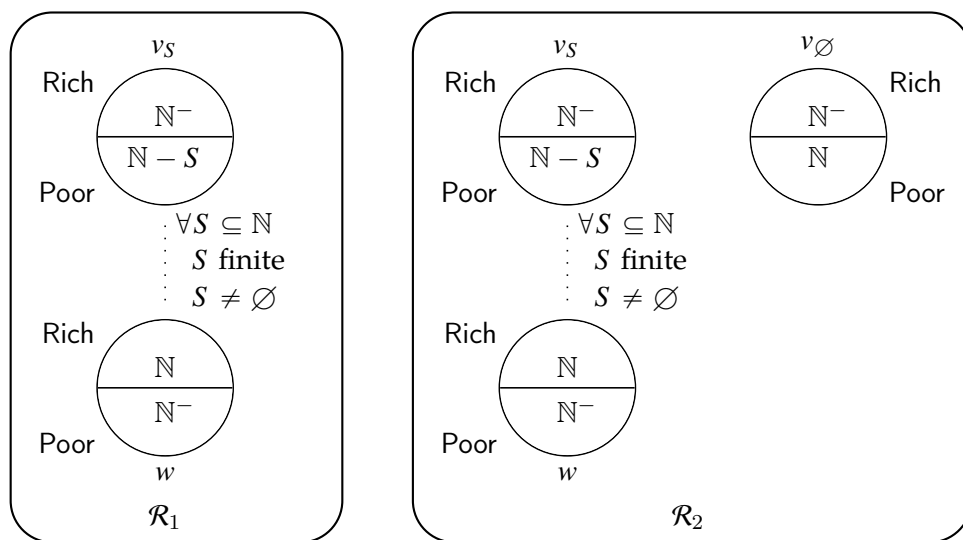
Figure 3: $\mathcal{L}^{1M}(@)$-bisimilar models disagreeing on (R). The top half of each circle satisfies Rich, while the bottom half satisfies Poor; at each v_S, the members of S do not exist.

Proof *(Sketch)*: Our game starts at $\mathcal{R}_1, w, w$ and $\mathcal{R}_2, w, w$. As before, we will describe a winning strategy for Eloïse such that, at every stage $\langle w, v, \overline{a}; w, v', \overline{b} \rangle$ of the $\mathcal{L}^{1M}(\approx, @)$-bisimulation game, $\mathcal{R}_1, w, v, \overline{a} \simeq \mathcal{R}_2, w, v', \overline{b}$.

Again, vacuously, $\mathcal{R}_1, w, w \simeq \mathcal{R}_2, w, w$. So suppose $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle$ is the current stage of the game, where $\mathcal{R}_1, w, u_1, \overline{a} \simeq \mathcal{R}_2, w, u_2, \overline{b}$ and where a_i is positive (i.e., in $\mathbb{N}$) iff b_i is positive. We will show that this continues to be true regardless of Abelard's move. If Abelard decides to pick a new object from $\delta_1(u_1)$ or $\delta_2(u_2)$, it will either be positive or negative. Since there are infinitely many of both, Eloïse will have no trouble picking a new one; and since there was a partial isomorphism between u_1 and u_2, Eloïse only needs the new objects to agree on their sign.

Suppose instead that Abelard decides to relocate the game. If he decides to move the game in both models back to w, since a_i is positive iff b_i is positive, we will have $\mathcal{R}_1, w, w, \overline{a} \simeq \mathcal{R}_2, w, w, \overline{b}$. (Likewise, if Abelard chooses to relocate to w in one model but lets Eloïse choose the other new location, she should still choose w for the reason above.) If he decides to relocate to some v_S where $S \neq \varnothing$ in, say, $\mathcal{R}_1$, let T be any set with the same cardinality as S such that $a_i \in S$ iff $b_i \in T$. Since there are only finitely many a_is, there will always be such a T. Eloïse can choose to relocate to v_T, and again, since a_i is positive iff b_i is positive, $\mathcal{R}_1, w, v_S, \overline{a} \simeq \mathcal{R}_2, w, v_T, \overline{b}$. Likewise if Abelard chooses to relocate to some v_S in $\mathcal{R}_2$.

The tricky part is determining what to do when Abelard decides to relocate to $v_\varnothing$ in $\mathcal{R}_2$. But since there are only finitely many a_is, Eloïse can just choose a v_S where $S \cap \{\overline{a}\} = \varnothing$. Then it will still be the case that $\mathcal{R}_1, w, v_S, \overline{a} \simeq \mathcal{R}_2, w, v_\varnothing, \overline{b}$. So no matter where Abelard decides to relocate, Eloïse can continue the game. ∎

Notice that the proof fails if we try to add either Π, $\downarrow$, or $\mathcal{F}$. It is easy to see this for Π, since we can express (4) as (2):

$$\Diamond \Pi x \, (@\mathsf{Rich}(x) \to \mathsf{Poor}(x)). \tag{2}$$

To see where the proof above fails for $\mathcal{L}^{1\mathsf{M}}(\approx, @, \Pi)$, consider what happens when Abelard decides to move from u_2 to $v_\varnothing$ in $\mathcal{R}_2$. Eloïse will try to match that move in $\mathcal{R}_1$ by moving from u_1 to some v_S where $S \cap \{\overline{a}\} = \varnothing$. But now, because of Π, Abelard is free to pick any object in S (and hence not in $\delta_1(v_S)$), forcing Eloïse to match it with an object in $\delta_2(v_\varnothing)$ (since $\delta_2(v_\varnothing) = D_2 = \mathbb{Z}$), and hence violating (Atomic).

As for $\downarrow$ and $\mathcal{F}$, the models above disagree on both of these formulas:

$$\exists x \, (\mathsf{Rich}(x) \wedge \Diamond \downarrow (\mathsf{Poor}(x) \wedge \Box \forall y \, @\mathsf{E}(y))) \tag{8}$$

$$\exists x \, (\mathsf{Rich}(x) \wedge \langle \mathcal{F} \rangle @ \, (\mathsf{Poor}(x) \wedge \Box \forall y \, @\mathsf{E}(y))). \tag{9}$$

In particular, $\mathcal{R}_2, w, w \models$ (8) and $\mathcal{R}_2, w, w \models$ (9) (take $v_\varnothing$ to be the world we shift to by $\Diamond \downarrow$ or $\langle \mathcal{F} \rangle @$), but $\mathcal{R}_1, w, w \not\models$ (8) and $\mathcal{R}_1, w, w \not\models$ (9). While this does not show that (4) can be expressed as an $\mathcal{L}^{1\mathsf{M}}(\approx, @, \downarrow, \mathcal{F})$-formula, it does show $\mathcal{R}_1$ and $\mathcal{R}_2$ cannot be used to settle the matter. Using modified models, however, it is possible to settle the matter in the negative: (4) is not even **UD**-expressible in $\mathcal{L}^{1\mathsf{M}}(\approx, @, \downarrow, \mathcal{F})$ (see §C).[23]

As a final example, we will show that even $\mathcal{L}^{1\mathsf{M}}(\approx, @, \Pi)$ cannot express (NR), which is formalized as (5):

$$\forall s \, (\mathsf{R}(s^*, s) \to \exists t \, (\mathsf{R}(s, t) \wedge \forall x \, (\mathsf{Rich}(x; s) \to \mathsf{Poor}(x; t)))). \tag{5}$$

Consider two models $\mathcal{N}_1 = \langle W_1, R_1, F_1, D_1, \delta_1, I_1 \rangle$ and $\mathcal{N}_2 = \langle W_2, R_2, F_2, D_2, \delta_2, I_2 \rangle$. Again, $D_1 = D_2 = \mathbb{Z}$ and the accessibility relations are universal. However, now all of $\mathbb{Z}$ exists at every world in either model. Our actual world is z, an egalitarian world where no integer is either rich or poor. For every finite set $S \subseteq \mathbb{N}$ and finite set $T \subseteq \mathbb{N}^-$, there is a world w_S^T where all the integers in $(\mathbb{N} - S) \cup T$ are rich, while all the integers in $(\mathbb{N}^- - T) \cup S$ are poor (so our old w is now just $w_\varnothing^\varnothing$). And for every *nonempty* finite set $S \subseteq \mathbb{N}$, and every finite set $T \subseteq \mathbb{N}^-$, there is a world v_S^T like before, where the rich and poor are flipped with respect to w_S^T. The only difference between $\mathcal{N}_1$ and $\mathcal{N}_2$ is the presence of worlds of the form $v_\varnothing^T$ in $\mathcal{N}_2$. See Figure 4 for a picture.[24]

$\mathcal{N}_1, z, z$ and $\mathcal{N}_2, z, z$ both agree that (4) is true, since nothing in z is rich. But they disagree on whether (5) is true; without the presence of $v_\varnothing^T$, there is no world where everyone rich in $w_\varnothing^\varnothing$ (i.e., $\mathbb{N}$) is poor. Thus, $\mathcal{N}_1 \not\models$ (5)$[z]$ but $\mathcal{N}_2 \models$ (5)$[z]$. Furthermore:

Proposition 4.4 (*Inexpressibility of (NR)*). $\mathcal{N}_1, z, z \leftrightarroweq_{\approx, @, \Pi} \mathcal{N}_2, z, z$. But $\mathcal{N}_2 \models$ (5)$[z]$ while $\mathcal{N}_1 \not\models$ (5)$[z]$. Hence, (5) is not expressible in $\mathcal{L}^{1\mathsf{M}}(\approx, @, \Pi)$.

---

[23]I claimed in Kocurek [2015, p. 215] that the proof of **Proposition 4.3** extends to $\mathcal{L}^{1\mathsf{M}}(\approx, @, \downarrow, \mathcal{F})$. This needs qualification. We can obtain a quick proof that (4) is not expressible in $\mathcal{L}^{1\mathsf{M}}(\approx, @, \downarrow, \mathcal{F})$ by restricting the accessibility relations in $\mathcal{R}_1$ and $\mathcal{R}_2$; but as §C reveals, the proof of **UD**-inexpressibility is more challenging.

[24]I claimed to prove this in Kocurek [2015, pp. 215–216] using models like the ones presented here, except with $T = \varnothing$ for all worlds. However, my proof was incorrect, since those models are distinguishable by the $\mathcal{L}^{1\mathsf{M}}$-formula $\Diamond \, [\exists x \, \mathsf{Rich}(x) \wedge \forall x \forall y \, (\mathsf{Rich}(x) \wedge \mathsf{Rich}(y) \to \Box (\mathsf{Rich}(x) \leftrightarrow \mathsf{Rich}(y)))]$.
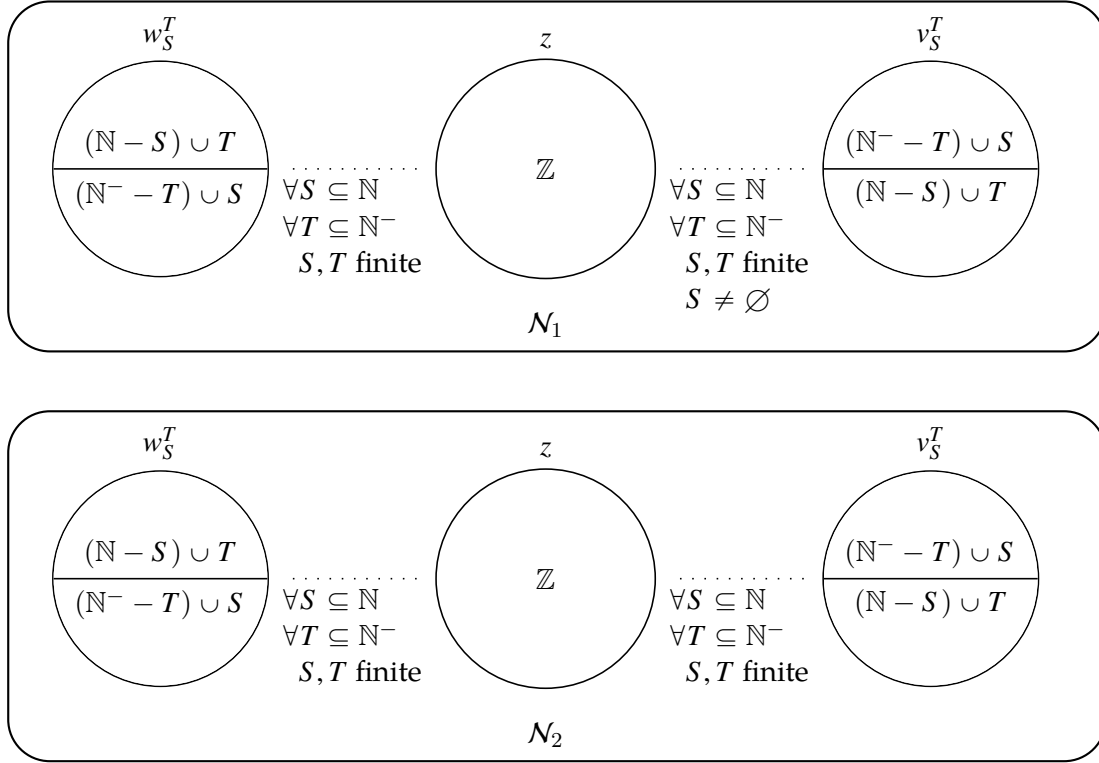
Figure 4: $\mathcal{L}^{1M}(\approx, @, \Pi)$-bisimilar models disagreeing on (NR).

Proof *(Sketch)*: Our game starts at $\mathcal{N}_1, z, z$ and $\mathcal{N}_2, z, z$. We will describe a winning strategy for Eloïse such that at every stage of the game $\langle z, u_1, \overline{a}; z, u_2, \overline{b} \rangle$, not only do we have $\mathcal{N}_1, z, u_1, \overline{a} \simeq \mathcal{N}_2, u_2, \overline{b}$, but also we have $u_1 = z$ iff $u_2 = z$, and we have $u_1 = w_{S_1}^{T_1}$ for some S_1 and T_1 iff $u_2 = w_{S_2}^{T_2}$ for some S_2 and T_2. Clearly this holds for the initial stage $\langle z, z; z, z \rangle$. So suppose $\langle z, u_1, \overline{a}; z, u_2, \overline{b} \rangle$ is such a stage.

Because for each world u and each $k \in \{1, 2\}$, $I_k(\text{Rich}, u)$ and $I_k(\text{Poor}, u)$ are empty or infinite, the back-and-forth step is easy. So suppose Abelard decides to relocate the game in $\mathcal{N}_1$. If he relocates to z, then Eloïse should also just relocate to z in $\mathcal{N}_1$. Suppose now Abelard decides to relocate to some $w_{S_1}^{T_1}$ in $\mathcal{N}_1$. Define $S_2 := \{ b_i \in \mathbb{N} \mid a_i \in I_1(\text{Poor}, w_{S_1}^{T_1}) \}$ and $T_2 := \{ b_i \in \mathbb{N}^- \mid a_i \in I_1(\text{Rich}, w_{S_1}^{T_1}) \}$. Then one can check that $\mathcal{N}_1, z, w_{S_1}^{T_1}, \overline{a} \simeq \mathcal{N}_2, z, w_{S_2}^{T_2}, \overline{b}$. For instance, suppose $a_i \in I_1(\text{Rich}, w_{S_1}^{T_1})$. Either $b_i \in \mathbb{N}$ or $b_i \in \mathbb{N}^-$. If $b_i \in \mathbb{N}$, then $b_i \notin S_2$, so $b_i \in I_2(\text{Rich}, w_{S_2}^{T_2})$. If $b_i \in \mathbb{N}^-$, then $b_i \in T_2$, so $b_i \in I_2(\text{Rich}, w_{S_2}^{T_2})$. Likewise, if $a_i \in I_1(\text{Poor}, w_{S_1}^{T_1})$, then $b_i \in I_2(\text{Poor}, w_{S_2}^{T_2})$. The same strategy works if Abelard chooses to relocate to some $v_{S_1}^{T_1}$ in $\mathcal{N}_1$. It also works if Abelard decides to relocate in $\mathcal{N}_2$, except when he chooses $v_{S_2}^{T_2}$ where the corresponding S_1 as defined above would be empty. In that case, for no $a_i \in \mathbb{N}$ is $b_i \in I_2(\text{Rich}, v_{S_2}^{T_2})$. So Eloïse can choose S_1 to be any nonempty subset of $\mathbb{N} - \{\overline{a}\}$. ∎

Once again, however, this inexpressibility proof does not extend to languages with ↓, since we can express (5) as:

$$\Box\downarrow\Diamond\Pi x\,(@\mathsf{Rich}(x) \to \mathsf{Poor}(x)).\tag{10}$$

Likewise, if we restrict to the class of models where $R = F$, we can express (5) with:

$$\mathcal{F}@\Diamond\Pi x\,(@\mathsf{Rich}(x) \to \mathsf{Poor}(x)).\tag{11}$$

§5   Generalizations

In the previous section, we saw a number of examples demonstrating how bisimulations can be used to prove inexpressibility results for a variety of two-dimensional logics. In this section, we turn to some more general results regarding the expressive power of two-dimensional modal languages and beyond.

First, given that sentences like (E), (R), and (NR) are expressible in some languages but not others, it is natural to ask what exactly the relative expressive power of all these various languages are. For instance, combining the results in §4, we know that $\mathcal{L}^{1M}(\approx) < \mathcal{L}^{1M}(\approx,@) < \mathcal{L}^{1M}(\approx,@,\Pi)$. But how do languages like $\mathcal{L}^{1M}(\approx,@,\downarrow)$ and $\mathcal{L}^{1M}(\approx,@,\mathcal{F})$ compare? Is one stronger than the other? What if we add a possibilist quantifier to one or the other?

Using bisimulation techniques like the ones in §4, we can mostly settle these questions for the two-dimensional languages discussed in this paper (though in what follows, I have excluded languages that include $\forall_@$). The inclusions relative to the class of all models can be diagrammed as in Figure 5 (for a proof of the accuracy of the diagram, see §D). The diagram is still accurate relative to **D**. Relative to **U**, adding ↓ or $\mathcal{F}$ without @ present is redundant. Moreover, $\mathcal{L}^{1M}(\mathsf{E},@,\downarrow,\Pi) \equiv_{\mathbf{U}} \mathcal{L}^{1M}(\mathsf{E},@,\mathcal{F},\Pi) \equiv_{\mathbf{U}} \mathcal{L}^{1M}(\mathsf{E},@,\downarrow,\mathcal{F},\Pi)$.[25] Relative to **UD**, we will have more inclusions. For instance, while $\mathcal{L}^{1M}(\Pi) \not\leq_{\mathbf{U}} \mathcal{L}^{1M}(@,\downarrow)$ and $\mathcal{L}^{1M}(\Pi) \not\leq_{\mathbf{D}} \mathcal{L}^{1M}(@,\downarrow)$, we do have $\mathcal{L}^{1M}(\Pi) \leq_{\mathbf{UD}} \mathcal{L}^{1M}(@,\downarrow)$, using the translation $\Pi x\varphi := \downarrow\Box\forall x@\varphi$. Similarly, $\mathcal{L}^{1M}(\Pi) \leq_{\mathbf{UD}} \mathcal{L}^{1M}(@,\mathcal{F})$. However, there are still limitations: $\mathcal{L}^{1M}(\Pi) \not\leq_{\mathbf{UD}} \mathcal{L}^{1M}(\approx,@)$ and $\mathcal{L}^{1M}(@,\Pi) \not\leq_{\mathbf{UD}} \mathcal{L}^{1M}(\approx,@,\downarrow,\mathcal{F})$. For more details, see §D.

Second, all of these inexpressibility results carry over to temporal logic. In temporal logic, one also includes a backward-looking operator $\Box^{-1}$, in addition to $\Box$, with the following semantic clause (where $R^{-1} = \{\langle v, w\rangle \mid \langle w, v\rangle \in R\}$):

$$\mathcal{M}, w, v, g \Vdash \Box^{-1}\varphi \quad\Leftrightarrow\quad \forall v' \in R^{-1}[v]\colon \mathcal{M}, w, v', g \Vdash \varphi.$$

Usually, $\Box$ and $\Box^{-1}$ are written respectively as G and H (for "it is always going to be" and "it has always been"), @ is written as N (for "now"), and ↓ is written as T (for "then"). The notion of a bisimulation can easily be generalized to temporal logic by including Zig-Zag clauses for both R and R^{-1} (which are often written respectively as $<$ and $>$).

All of the sentences considered in this paper have natural temporal analogues. Here are a few variations on some of the sentences we have been considering (where R is replaced by $<$):

---

[25]The exact relation between $\mathcal{L}^{1M}(@,\mathcal{F})$, $\mathcal{L}^{1M}(@,\downarrow)$, and $\mathcal{L}^{1M}(@,\downarrow,\mathcal{F})$ relative to **U** is an open question. In particular, it is unknown whether $\mathcal{L}^{1M}(@,\downarrow) <_{\mathbf{U}} \mathcal{L}^{1M}(@,\mathcal{F})$ or even $\mathcal{L}^{1M}(@,\downarrow,\mathcal{F}) \equiv_{\mathbf{U}} \mathcal{L}^{1M}(@,\mathcal{F})$. Figure 11 in §D summarizes the remaining inclusions relative to **U**.
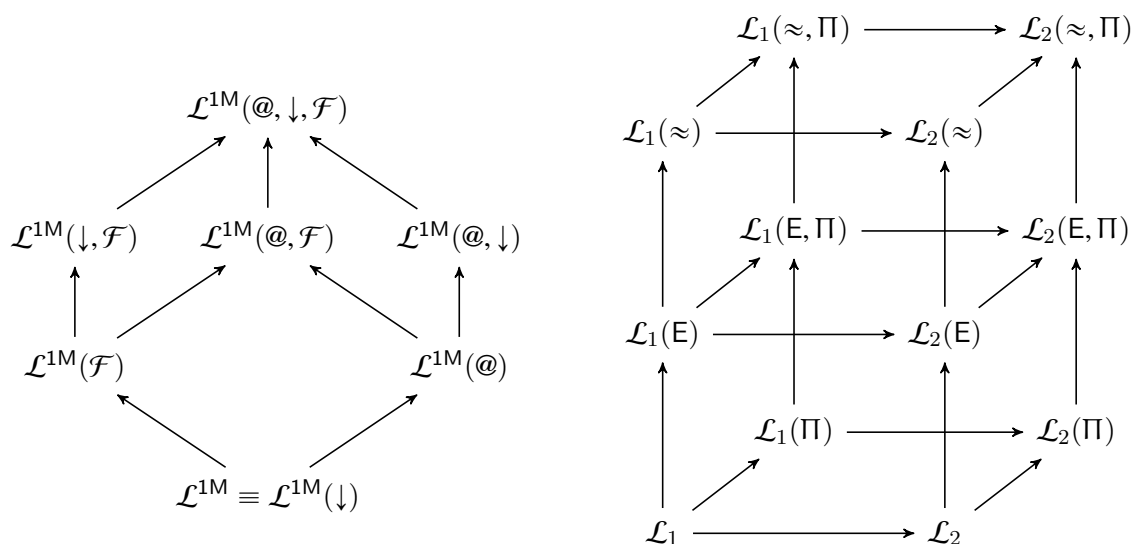
Figure 5: The (**D**-)expressive hierarchy for two-dimensional languages between $\mathcal{L}^{1M}$ and $\mathcal{L}^{1M}(\approx, @, \downarrow, \mathcal{F}, \Pi)$. Arrows represent strict increase in expressive power. If there is an upward path from $\mathcal{L}_1$ to $\mathcal{L}_2$ in the diagram on the left, then the inclusions between $\mathcal{L}_1$, $\mathcal{L}_2$, and their extensions with E, $\approx$, or Π are represented in the right diagram.

(FE) There will be things other than there are now.

$$\exists t > s^* \ (\mathsf{E}(x; t) \wedge \neg \mathsf{E}(x; s^*)) \tag{12}$$

(PR) It was the case that everyone now rich was poor.

$$\exists t < s^* \, \forall x \ (\mathsf{Rich}(x; s^*) \rightarrow \mathsf{Poor}(x; t)) \tag{13}$$

(FPR) Henceforth, everyone who is rich will have once been poor.

$$\forall t > s^* \, \exists t' < t \, \forall x \ (\mathsf{Rich}(x; t) \rightarrow \mathsf{Poor}(x; t')) \tag{14}$$

All of our models in §4 have universal accessibility relations. However, allowing $<$ to be universal would be too permissive in the context of temporal logic (there would be no difference between future and past!). Often, $<$ is required to be at least a strict partial order (i.e., irreflexive, asymmetric, and transitive), thereby excluding models where it is universal. Thus, none of the results in §4 immediately carry over to temporal logic.

Fortunately, we can still piggyback on these inexpressibility results with the following trick. Suppose $\mathcal{M}, w, w \leftrightarroweq_{S_1,\dots,S_n} \mathcal{N}, v, v$ where the accessibility relations of $\mathcal{M}$ and $\mathcal{N}$ are universal. Assume for simplicity that $W^{\mathcal{M}}$ and $W^{\mathcal{N}}$ are countable. Let $f^{\mathcal{M}} : \mathbb{N} \to W^{\mathcal{M}}$ be a surjection such that $f^{\mathcal{M}}(0) = w$ (and likewise for $f^{\mathcal{N}}$). Define a new model $\mathcal{M}_{\mathbb{Z} \times \mathbb{N}}$ where $W^{\mathcal{M}_{\mathbb{Z} \times \mathbb{N}}} := \mathbb{Z} \times W^{\mathcal{M}}$, $R^{\mathcal{M}_{\mathbb{Z} \times \mathbb{N}}} = F^{\mathcal{M}_{\mathbb{Z} \times \mathbb{N}}} := \{\langle \langle i, f(n) \rangle, \langle j, f(m) \rangle \rangle \mid i < j \text{ or } (i = j \text{ and } n < m)\}$, $D^{\mathcal{M}_{\mathbb{Z} \times \mathbb{N}}} := D^{\mathcal{M}}$, $\delta^{\mathcal{M}_{\mathbb{Z} \times \mathbb{N}}}(\langle i, f(n) \rangle) := \delta^{\mathcal{M}}(f(n))$, and $I^{\mathcal{M}_{\mathbb{Z} \times \mathbb{N}}}(P, \langle i, f(n) \rangle) := I^{\mathcal{M}}(P, f(n))$. That is,

each $i \in \mathbb{Z}$ contains a copy of $\mathcal{M}$, and the integer-world pairs are ordered lexicographically. Define $\mathcal{N}_{\mathbb{Z} \times \mathbb{N}}$ similarly from $\mathcal{N}$ using $f^{\mathcal{N}}$.

It is straightforward to check that $\mathcal{M}_{\mathbb{Z} \times \mathbb{N}}, \langle 0, w \rangle, \langle 0, w \rangle \leftrightarrows_{\square^{-1}, S_1, \ldots, S_n} \mathcal{N}_{\mathbb{Z} \times \mathbb{N}}, \langle 0, v \rangle, \langle 0, v \rangle$. For instance, suppose we are playing the $\mathcal{L}^{1M}(\square^{-1}, S_1, \ldots, S_n)$ back-and-forth game with $\mathcal{M}_{\mathbb{Z} \times \mathbb{N}}$ and $\mathcal{N}_{\mathbb{Z} \times \mathbb{N}}$, and we are currently at stage $\langle \langle i_1, u_1 \rangle, \langle i_2, u_2 \rangle, \overline{a}; \langle i'_1, u'_1 \rangle, \langle i'_2, u'_2 \rangle, \overline{b} \rangle$. Then whenever Abelard makes a move, Eloïse need only consult the $\mathcal{L}^{1M}(S_1, \ldots, S_n)$ back-and-forth with $\mathcal{M}$ and $\mathcal{N}$, and see how she would respond at stage $\langle u_1, u_2, \overline{a}; u'_1, u'_2, \overline{b} \rangle$. In particular, if Abelard chooses to relocate the game in $\mathcal{M}_{\mathbb{Z} \times \mathbb{N}}$ to $\langle i_3, u_3 \rangle$ where $\langle i_3, u_3 \rangle > \langle i_2, u_2 \rangle$, then Eloïse can pick whatever u'_3 she would have chosen had Abelard chose u_3 in the back-and-forth game with $\mathcal{M}$ and $\mathcal{N}$, and then she can relocate the game in $\mathcal{N}_{\mathbb{Z} \times \mathbb{N}}$ to $\langle i'_3, u'_3 \rangle$ where $\langle i'_3, u'_3 \rangle > \langle i'_2, u'_2 \rangle$ (she will always be able to find one, since there are infinitely many copies of $\mathcal{N}$ after i'_2). The same strategy applies if Abelard picks $\langle i_3, u_3 \rangle$ with $\langle i_3, u_3 \rangle < \langle i_2, u_2 \rangle$. Thus, our results in §4 can be extended to temporal logic.[26]

Finally, it seems as though two-dimensions is not enough to overcome the kind of expressive limitations discussed in this paper in full generality. Recall that while (NR) is not expressible as an $\mathcal{L}^{1M}(\approx, @, \Pi)$-formula, it is expressible as an $\mathcal{L}^{1M}(\approx, @, \downarrow, \Pi)$-formula. However, more complicated sentences can be constructed that reveal the expressive limitations of even $\mathcal{L}^{1M}(\approx, @, \downarrow, \Pi)$. The most natural examples use temporal modalities or mix modalities. For instance, here is a temporal example from Cresswell [1990, p. 29]:

(H)   Once, everyone now alive who was not then miserable would eventually be happy.

To formalize this, it seems that we need to store two reference times, not just one. In $\mathcal{L}^{TS}$, we would formalize (H) as follows:

$$\exists t < s^* \, \exists t' > t \, \forall x \, ((\mathsf{Alive}(x; s^*) \wedge \neg \, \mathsf{Miserable(x;t)}) \rightarrow \mathsf{Happy}(x; t')) . \qquad (15)$$

We can also get examples with metaphysical modality, although the more powerful the language is, the more contrived the examples have to be:

(RC)   There could have been a brave man such that everyone who was poor but kind in reality necessarily received money from that man.

The problem is that we need to be able to go back to both the actual world and the first world we shifted to while we are at the second world we shifted to; but we can only keep track of one reference world at a time. It has been noted in the literature that this point seems to generalize to higher-dimensional languages.[27] One gets the feeling that for any n, we can concoct a $\mathcal{L}^{TS}$-formula that requires keeping track of $(n + 1)$-many worlds in our points of evaluation. But no proof of this claim has been offered in the literature.[28]

---

[26]Though our partial order was linear and discrete, this was not crucial to the construction. We could have, for instance, mapped each state in $\mathcal{M}$ to a rational in the interval $[0, 1)$, and have obtained a dense linear order. Alternatively, via tree unraveling, we could have obtained a branching structure.

[27]See, e.g., Vlach [1973, p. 183–185], Needham [1975, pp. 73–74], van Benthem [1977, p. 418], Forbes [1989, p. 87], and Cresswell [1990, pp. 29–30].

[28]Several conjectures have been made about how to construct such formulas. For instance, Needham [1975, pp. 73–74] gives a sentence he claims is not expressible in $\mathcal{L}^{1M}(\square^{-1}, @, \downarrow, \Pi)$; but van Benthem [1977, p. 417] shows it is. van Benthem then gives a genuine example of a temporal $\mathcal{L}^{TS}$-formula that is not expressible

Using the power of bisimulations, we can actually verify this claim. I will conclude by explaining how to generate further inexpressibility results for higher-dimensional languages. We will only explicitly prove that a certain three-dimensional language is not expressible in any two-dimensional language. Hopefully, it will be clear how the method can be schematized to show that some $(n+1)$-dimensional languages are not expressible in any n-dimensional language.

First, an ***n-dimensional model*** is a tuple of the form $\mathcal{M} = \langle W, R, R_1, \dots, R_{n-1}, D, \delta, I \rangle$ where each $R_i \subseteq W \times W$, and otherwise everything is as before. For instance, the models we have been working with in this paper have all been 2-dimensional models (where $F = R_1$). Second, for each $k \geqslant 1$, we will introduce operators $\square_k$, $@_k$, and $\downarrow_k$, which we might add to $\mathcal{L}^{1M}$. We will define $\mathcal{L}_1^{1M} := \mathcal{L}^{1M}$ and $\mathcal{L}_{n+1}^{1M} := \mathcal{L}^{1M}(\square_1, \dots, \square_n, @_1, \dots, @_n, \downarrow_1, \dots, \downarrow_n)$. Third, where $\sigma = \langle w_1, \dots, w_n \rangle$ is a sequence of worlds, and where $1 \leqslant k \leqslant n$, let σ_v^k be the result of replacing w_k in σ with v. Finally, satisfaction for $\mathcal{L}_{n+1}^{1M}$ will be relativized to $(n+1)$-dimensional models, as well as a sequence of n-many worlds $\sigma = \langle w_1, \dots, w_n \rangle$, a world v, and a variable assignment $g \in \mathsf{VA}(\mathcal{M})$, with the semantic clauses for the new operators stated below for $1 \leqslant k \leqslant n$:

$$\mathcal{M}, \sigma, v, g \Vdash \square_k \varphi \quad \Leftrightarrow \quad \forall u \in R_k[w_k] \colon \mathcal{M}, \sigma_u^k, v, g \Vdash \varphi$$

$$\mathcal{M}, \sigma, v, g \Vdash @_k \varphi \quad \Leftrightarrow \quad \mathcal{M}, \sigma, w_k, g \Vdash \varphi$$

$$\mathcal{M}, \sigma, v, g \Vdash {\downarrow_k} \varphi \quad \Leftrightarrow \quad \mathcal{M}, \sigma_v^k, v, g \Vdash \varphi.$$

Thus, $\mathcal{F}$, $@$, and $\downarrow$ are $\square_1$, $@_1$, and $\downarrow_1$ respectively. Since $\mathcal{L}_1^{1M} = \mathcal{L}^{1M}$ and $\mathcal{L}_2^{1M}$ is essentially $\mathcal{L}^{1M}(@, \downarrow, \mathcal{F})$, it makes sense to call $\mathcal{L}_n^{1M}$ with this semantics an ***n-dimensional language***. Generalizing the definition of a bisimulation to $\mathcal{L}_n^{1M}$ is straightforward.

Using models similar to $\mathcal{N}_1$ and $\mathcal{N}_2$ from Figure 4, we can show that $\mathcal{L}_{n+1}^{1M}$ is not included in $\mathcal{L}_n^{1M}(\approx, \Pi)$. We have already shown with **Proposition 4.4** that $\mathcal{L}_2^{1M}$ is not included in $\mathcal{L}_1^{1M}(\approx, \Pi)$ (or even in $\mathcal{L}_1^{1M}(\approx, @_1, \Pi)$), since $\square{\downarrow_1}\Diamond \forall x \, (@_1\mathsf{Rich}(x) \to \mathsf{Poor}(x))$ distinguishes $\mathcal{N}_1, z, z$ and $\mathcal{N}_2, z, z$, even though $\mathcal{N}_1, z, z \leftrightarroweq_{\approx, @_1, \Pi} \mathcal{N}_2, z, z$. We will show that the following $\mathcal{L}_3^{1M}$-formula is not expressible in $\mathcal{L}_2^{1M}(\approx, \Pi)$:[29]

$$\square{\downarrow_1}\square{\downarrow_2}\Diamond \forall x \, \left((@_1 P_1(x) \wedge @_2 P_2(x)) \to P_3(x) \right). \tag{16}$$

---

in $\mathcal{L}^{1M}(\square^{-1}, @, \downarrow, \Pi)$. However, it should be noted that even though $\mathcal{F}$ operators were not the focus of van Benthem [1977], the sentence he gives is expressible in $\mathcal{L}^{1M}(@, \mathcal{F}^{-1}, \Pi)$, which is still two-dimensional. Forbes [1989, p. 89] also gives a schema that was supposed to show that $(n-1)$-dimensional logic is not as expressive as n-dimensional logic. But as footnote 29 explains, the example is not correct either. Cresswell [1990, p. 30] suggests one can generate such sentences from (H) since, reading the conditional in (15) as a disjunction, "disjunctions can be extended with no upper limit."

[29] Forbes [1989, p. 87] gives a purported example of an n-dimensional formula not expressible as an $(n+1)$-dimensional formulas. (Following Forbes, we restrict attention to models whose accessibility relations are universal.) Given a model $\mathcal{M}$, let us say a sequence $w_1, \dots, w_n \in W^{\mathcal{M}}$ is an ***n-chain*** if $\delta^{\mathcal{M}}(w_i) \subset \delta^{\mathcal{M}}(w_{i+1})$ for $1 \leqslant i < n$. Forbes [1989, p. 89] says "it is a very probable conjecture that 'there is an n-chain of worlds' cannot be expressed in $[\mathcal{L}_{n-1}^{1M}(\approx)]$, so that the hierarchy of modal languages is strictly increasing in expressive power…" Forbes also notes that $\mathcal{L}_{n-1}^{1M}(\approx, \Pi)$ can express "there is an n-chain of worlds", but claims that it cannot express "there is an $n+1$-chain of worlds". However, the conjecture is false. Define $\theta_{i,k} := @_i \forall x \, @_k \mathsf{E}(x) \wedge @_k \exists x \, @_i \neg \mathsf{E}(x)$. Then the claim that there is a 4-chain can be expressed as an $\mathcal{L}_3^{1M}(\approx)$-formula: $\Diamond_1 \Diamond_2 (\theta_{1,2} \wedge \Diamond_1 (\theta_{2,1} \wedge \Diamond_2 \theta_{1,2}))$. Likewise for $n > 4$. Also, let $\chi := \Pi x \, (@_1 \mathsf{E}(x) \to \mathsf{E}(x)) \wedge \exists x \, @_1 \neg \mathsf{E}(x)$ and let $\eta := \forall x \, @_1 \mathsf{E}(x) \wedge \Sigma x \, (@_1 \mathsf{E}(x) \wedge \neg \mathsf{E}(x))$. Then the claim that there is a 4-chain can be expressed as an $\mathcal{L}_2^{1M}(\approx, \Pi)$-formula: $\Diamond_1 \Diamond (\chi \wedge \Diamond_1 (\eta \wedge \Diamond \chi))$. Again, this generalizes to n-chains where $n > 4$.

21

The proof that $\Box\!\downarrow_1 \cdots \Box\!\downarrow_n \Diamond \forall x \ (\bigwedge_{i=1}^n @_i P_i(x) \to P_{n+1}(x))$ is not expressible in $\mathcal{L}_n^{1\mathsf{M}}(\approx, \Pi)$ is a straightforward generalization of the one below.

First, we describe our models $\mathcal{N}_3$ and $\mathcal{N}_4$. These models have been summarized in Figure 6. All of the accessibility relations will be universal, and $\mathcal{N}_3$ and $\mathcal{N}_4$ will both be constant domain models (so the local domain of every world is the global domain). Let $\mathbb{N}_1$, $\mathbb{N}_2$, $\mathbb{N}_3$, and $\mathbb{N}_4$ be disjoint copies of $\mathbb{N}$. In both models, there will be a unique world z where $I_k(P_1, z) = I_k(P_2, z) = I_k(P_3, z) = \varnothing$ for $k \in \{3, 4\}$. There will be three types of worlds: α-worlds, β-worlds, and γ-worlds. Each type of world will be uniquely specified by four sets S_1, S_2, S_3, and S_4, where $S_k \subseteq \mathbb{N}_k$ and $|S_k| < \aleph_0$ for $k \in \{1, 2, 3, 4\}$. Where $\eta \in \{\alpha, \beta, \gamma\}$, we will denote the worlds as $\eta_{S_1, S_2, S_3, S_4}$. Each of S_1, S_2, S_3, and S_4 is generally allowed to be empty, but in $\mathcal{N}_3$, only γ-worlds where $S_1 \neq \varnothing$ are allowed. For each $k \in \{3, 4\}$ and each $i \in \{1, 2, 3\}$, $I_k(P_i, \eta_{S_1, S_2, S_3, S_4}) = \varnothing$ with the following exceptions:

- $I_k(P_1, \alpha_{S_1, S_2, S_3, S_4}) = (\mathbb{N}_1 - S_1) \cup (\mathbb{N}_2 - S_2) \cup S_3 \cup S_4$

- $I_k(P_2, \beta_{S_1, S_2, S_3, S_4}) = (\mathbb{N}_1 - S_1) \cup (\mathbb{N}_3 - S_3) \cup S_2 \cup S_4$

- $I_k(P_3, \gamma_{S_1, S_2, S_3, S_4}) = (\mathbb{N}_1 - S_1) \cup (\mathbb{N}_4 - S_4) \cup S_2 \cup S_3$.
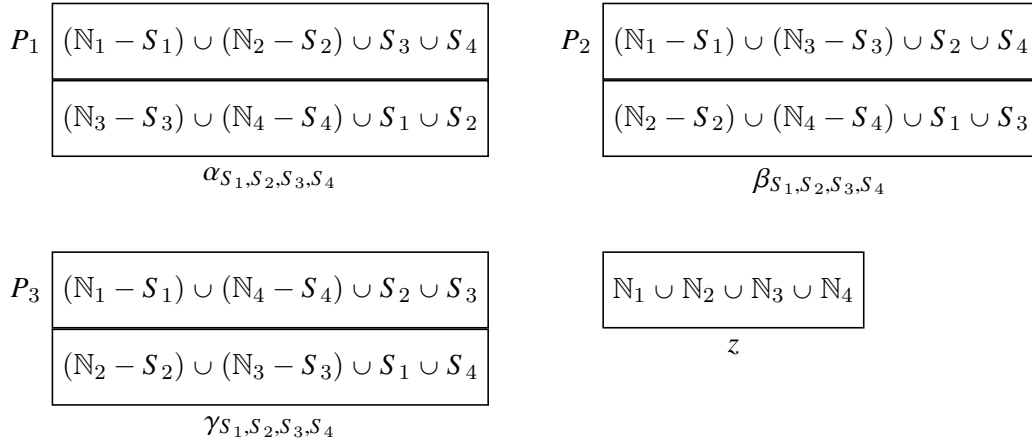

Figure 6: Summary of $\mathcal{N}_3$ and $\mathcal{N}_4$. For each $k \in \{1, 2, 3, 4\}$, we have that $S_k \subseteq \mathbb{N}_k$ and $|S_k| < \aleph_0$. For $\mathcal{N}_3$, there are no worlds of the form $\gamma_{\varnothing, S_2, S_3, S_4}$.

We will start by explaining why $\mathcal{N}_3, \langle z, z \rangle, z \not\Vdash (16)$ while $\mathcal{N}_4, \langle z, z \rangle, z \Vdash (16)$. First, to explain why $\mathcal{N}_3, \langle z, z \rangle, z \not\Vdash (16)$, it suffices to note the following (where α and β below have set $S_1 = S_2 = S_3 = S_4 = \varnothing$, and w is any world):

$$\mathcal{N}_3, \langle \alpha, \beta \rangle, w \not\Vdash \Diamond \forall x \ ((@_1 P_1(x) \wedge @_2 P_2(x)) \to P_3(x)).$$

This is simply because $I(P_1, \alpha) \cap I(P_2, \beta) = \mathbb{N}$, but no γ-world has all of $\mathbb{N}$ in its interpretation of P_3. Second, to explain why $\mathcal{N}_4, \langle z, z \rangle, z \Vdash (16)$, consider the following formula:

$$\Diamond \forall x \ ((@_1 P_1(x) \wedge @_2 P_2(x)) \to P_3(x)). \tag{17}$$

Notice that $\mathcal{N}_4, \langle w, v \rangle, u \Vdash$ (17) holds vacuously unless w is an α-world and v is a β-world. So suppose $w = \alpha_{S_1, S_2, S_3, S_4}$ and $v = \beta_{T_1, T_2, T_3, T_4}$. Then:

$$I_4(P_1, w) \cap I_4(P_2, v) = (\mathbb{N}_1 - (S_1 \cup T_1)) \cup (T_2 - S_2) \cup (S_3 - T_3) \cup (S_4 \cap T_4).$$

Now pick $u = \gamma_{(S_1 \cup T_1), (T_2 - S_2), (S_3 - T_3), S_4'}$, where S_4' is disjoint from $S_4 \cap T_4$. (We can always find such a γ since $S_1 \cup T_1$ is allowed to be empty.) Then it is easy to see that $\mathcal{N}_4, \langle w, v \rangle, u \Vdash \forall x \, ((@_1 P_1(x) \wedge @_2 P_2(x)) \to P_3(x))$. Thus, $\mathcal{N}_3, \langle z, z \rangle, z$ and $\mathcal{N}_4, \langle z, z \rangle, z$ disagree on (16).

Now, in what follows, let us say that a **2D-partial isomorphism** between $\mathcal{M}, \langle w, v \rangle, v, \overline{a}$ and $\mathcal{N}, \langle w', z' \rangle, v', \overline{b}$ is a map ρ such that ρ is a partial isomorphism between $\mathcal{M}, w, v, \overline{a}$ and $\mathcal{N}, w', v', \overline{b}$, and also a partial isomorphism between $\mathcal{M}, w, w, \overline{a}$ and $\mathcal{N}, w', w', \overline{b}$. In other words, 2D-partial isomorphisms must also satisfy (Predicate) and (Existence) at w and w'. We will write $\simeq_{2D}$ in place of $\simeq$ for 2D-partial isomorphisms. It is easy to verify that 2D-partial isomorphism allows for Eloïse to continue the $\mathcal{L}_2^{1M}(\approx)$-bisimulation game.

Let us write $\mathcal{M}, w, z, v, \overline{a} \leftrightarroweq^2 \mathcal{N}, w', z', v', \overline{b}$ to mean that $\mathcal{M}, w, z, u, \overline{a}$ and $\mathcal{N}, w', z', u', \overline{b}$ are $\mathcal{L}_2^{1M}$-bisimilar (I am dropping the angle brackets). Note $\leftrightarroweq^2$ is just $\mathcal{L}_2^{1M}(@_1, \downarrow_1, \Box_1) = \mathcal{L}^{1M}(@, \downarrow, \mathcal{F})$-bisimilarity, with an extra argument place for worlds in the middle; no clause in this notion of bisimilarity can do anything to or with the z and z' worlds. The conventions from before regarding adding additional operators, quantifiers, etc. all apply.

Theorem 5.1 (*Higher-Dimensional Inexpressibility*). $\mathcal{N}_3, z, z, z \leftrightarroweq^2_{\approx, @_2, \Pi} \mathcal{N}_4, z, z, z$.

Proof: Clearly, $z, z, z \simeq_{2D} z, z, z$. Now, suppose $w, z, v, \overline{a} \simeq_{2D} w', z, v', \overline{b}$, where:

- $w = z$ iff $w' = z$ (likewise for v and v')

- w is an $\alpha/\beta/\gamma$-world iff w' is an $\alpha/\beta/\gamma$-world (likewise for v and v')

- $w = v$ iff $w' = v'$.

Using these assumptions, we will show that no matter what move Abelard makes, Eloïse can match his move so as to preserve these assumptions.

First, observe that no matter what worlds w and v are, for any $i, j \in \{1, 2, 3\}$, all of the following sets are either empty or infinite (where $k \in \{3, 4\}$):

$$I_k(P_i, w) \cap I_k(P_j, v) \qquad \overline{I_k(P_i, w)} \cap I_k(P_j, v)$$
$$I_k(P_i, w) \cap \overline{I_k(P_j, v)} \qquad \overline{I_k(P_i, w)} \cap \overline{I_k(P_j, v)}.$$

Thus, if Abelard picks a new $a \in \delta_3(u)$, then no matter what predicates it satisfies in w or v, Eloïse will have infinitely many new $b \in \delta_4(u')$ to choose from which satisfy the same predicates in w' and v'. Likewise if Abelard picks a new $b \in \delta_4(u')$.

Next, suppose Abelard decides to relocate the game in $\mathcal{N}_3$. Since this is the $\mathcal{L}_2^{1M}(\approx, \Pi)$-bisimulation game, he can either choose to replace w with another world, or v with another world (there are no clauses for resetting z). Clearly if he replaces one of these with z, Eloïse should match his move by replacing the corresponding

world with z. If he replaces w with v, Eloïse should replace w' with v'. Likewise if he replaces v with w. Suppose WLOG that he decides to replace v with a new α-world $u = \alpha_{S_1, S_2, S_3, S_4}$. Define:

$$T_1 := \{b_i \in \mathbb{N}_1 \mid a_i \notin I_3(P_1, u)\} \qquad T_3 := \{b_i \in \mathbb{N}_3 \mid a_i \in I_3(P_1, u)\}$$
$$T_2 := \{b_i \in \mathbb{N}_2 \mid a_i \notin I_3(P_1, u)\} \qquad T_4 := \{b_i \in \mathbb{N}_4 \mid a_i \in I_3(P_1, u)\}.$$

Define $u' = \alpha_{T_1, T_2, T_3, T_4}$. It is easy to check that $a_i \in I_3(P_1, u)$ iff $b_i \in I_4(P_1, u')$. (Suppose $a_i \in I_3(P_1, u)$, and reason by cases depending on which $\mathbb{N}_k$ contains b_i. Likewise with $a_i \notin I_3(P_1, u)$.) The other cases are symmetric (for both Zig and Zag), except if Abelard decides to replace v' with a γ-world where the corresponding S_1 we define would be empty. In that case, define the other sets as before, and pick a new $c \in \mathbb{N}_1 - \{\overline{a}\}$ and set $S_1 = \{c\}$.   ∎

§6 Conclusion

As we have seen, there are a number of English modal claims that seem to resist regimentation in first-order modal logic, even if we add two-dimensional operators. Proofs of these claims in the literature are often quite complicated. But as we have shown, they can be simplified by first regimenting these English sentences as formulas in an extensional two-sorted language and then constructing bisimilar models that disagree on these extensional formulas. We illustrated this technique by showing that (E) is not expressible in $\mathcal{L}^{1M}(\approx)$, that (R) is not expressible in $\mathcal{L}^{1M}(\approx, @)$, and that (NR) is not expressible in $\mathcal{L}^{1M}(\approx, @, \Pi)$. We then classified the relative expressive power of the extensions of $\mathcal{L}^{1M}$ with operators like $@$, $\downarrow$, and $\mathcal{F}$, and finally showed how these inexpressibility results generalize to temporal languages and to higher-dimensional languages.

There are still a number of questions about the expressivity of extensions of first-order modal logic that have yet to be resolved. For one thing, we have yet to complete the classification of the expressive power of these languages relative to **U** and **UD**, and one might also wonder whether these results hold in other kinds of models, such as the class of models with finite global domains.[30] More broadly, there is still a question about whether we can formally characterize sentences like (E), (R), and (NR).[31] Finally, there is still much to be learned about even more powerful extensions of first-order modal logic. For instance, the results in §5 suggest that the key to overcoming all of these expressive limitations is to move to a hybrid language, or some weaker Vlachian languages.[32] Bisimulation techniques can also be used to characterize the expressive power of first-order Vlachian logics

---

[30]A theorem of Yanovich [2015, pp. 85–86] shows that one will not generally be able to use the bisimulation technique proposed in this paper to establish inexpressibility over the class of models with finite domains. I suspect one could still establish such results, however, by constructing sequences of pairs of finite models that were n-bisimilar for each $n \in \mathbb{N}$. But working out the details must be left for future work.

[31]See Kocurek [2016] for one such syntactic characterization.

[32]See Vlach [1973]; Areces et al. [1999]; Areces and ten Cate [2007]; Fritz [2012]; Yanovich [2015]; Kocurek [2016].

and hybrid logics.[33] But there is still much to uncover about the full expressive landscape for these languages and their extensions.

§A   van Benthem's Characterization Theorem

In this appendix, we prove **Theorem 3.5**. We essentially follow the proof of the corresponding theorem for propositional modal logic in Blackburn et al. [2001, Chp. 2.6]. The crucial change is with the definition of a set of formulas being satisfiable.

Definition A.1 *(Satisfiability)*. Let $\bar{z}$ be some new variables not in VAR, and let $\mathcal{L}(\bar{z})$ be the result of extending $\mathcal{L}$ with $\bar{z}$. Let $\Gamma(\bar{x}, \bar{z})$ be a set of $\mathcal{L}(\bar{z})$-formulas whose only variables not among $\bar{z}$ are $\bar{x}$. Let $\mathcal{M}$ be a model, and $X \subseteq W^2$, and let $\bar{b} \in D$.

- Γ is $\Sigma/\exists/\exists_@$-**satisfiable in X over $\bar{b}$ (with respect to $\mathcal{M}$)** if there is a $\langle w, v \rangle \in X$ and some $\bar{a} \in D/\delta(v)/\delta(w)$ such that $\mathcal{M}, w, v \Vdash \Gamma[\bar{a}, \bar{b}]$.

- Γ is *finitely* $\Sigma/\exists/\exists_@$-**satisfiable in X over $\bar{b}$ (with respect to $\mathcal{M}$)** if every finite subset of Γ is $\Sigma/\exists/\exists_@$-satisfiable in X over $\bar{b}$ (with respect to $\mathcal{M}$).

Note that if the only free variables in Γ are among $\bar{z}$, then Γ is (finitely) Σ-satisfiable in X over $\bar{b}$ iff it is (finitely) $\exists$-satisfiable in X over $\bar{b}$ iff it is (finitely) $\exists_@$-satisfiable in X over $\bar{b}$. We will just use the term "(finitely) satisfiable" when the distinction does not matter.

Definition A.2 *(Modal Saturation)*. Assume $\mathcal{F}$, $\forall_@$, and Π are not among $S_1, \ldots, S_n$. A model $\mathcal{M}$ is $\mathcal{L}^{1M}(S_1, \ldots, S_n)$-**saturated** if for all $w, v \in W$, all $\bar{b} \in D$, and all sets $\Gamma(\bar{x}, \bar{z})$ of $\mathcal{L}^{1M}(S_1, \ldots, S_n, \bar{z})$-formulas:

(a)   if Γ is finitely $\exists$-satisfiable in $\{w\} \times R[v]$ over $\bar{b}$ (with respect to $\mathcal{M}$), then it is $\exists$-satisfiable in $\{w\} \times R[v]$ over $\bar{b}$;

(b)   if Γ is finitely $\exists$-satisfiable in $\{\langle w, v \rangle\}$ over $\bar{b}$ (with respect to $\mathcal{M}$), then it is $\exists$-satisfiable in $\{\langle w, v \rangle\}$ over $\bar{b}$.

If $\mathcal{F}$ is among $S_1, \ldots, S_n$, we just add the following clause:

(c)   if Γ is finitely $\exists$-satisfiable in $F[w] \times \{v\}$ over $\bar{b}$ (with respect to $\mathcal{M}$), then it is $\exists$-satisfiable in $F[w] \times \{v\}$ over $\bar{b}$.

If $\Pi/\forall_@$ is among $S_1, \ldots, S_n$, add the clauses above but with $\exists$ replaced by $\Sigma/\exists_@$.

Lemma A.3 (*Modal Saturation implies the Hennessy-Milner Property*). Suppose $\mathcal{M}$ and $\mathcal{N}$ are $\mathcal{L}^{1M}(S_1, \ldots, S_n)$-saturated. Then $\equiv_{S_1,\ldots,S_n}$ is an $\mathcal{L}^{1M}(S_1, \ldots, S_n)$-bisimulation between $\mathcal{M}$ and $\mathcal{N}$. Hence, if $\mathcal{M}, w, v, \bar{a} \equiv_{S_1,\ldots,S_n} \mathcal{N}, w', v', \bar{b}$, then it follows that $\mathcal{M}, w, v, \bar{a} \leftrightarroweq_{S_1,\ldots,S_n} \mathcal{N}, w', v', \bar{b}$.

Proof: Suppose $\mathcal{M}, w, v, \bar{a} \equiv_{S_1,\ldots,S_n} \mathcal{N}, w', v', \bar{b}$. Clearly (Atomic) is satisfied (and likewise for (Ex) and (Eq) if E or $\approx$ are among $S_1, \ldots, S_n$).

Zig-Zag. Let $u \in R^{\mathcal{M}}[v]$. Define $\Gamma(\bar{z}) := \{\varphi(\bar{z}) \mid \mathcal{M}, w, u \Vdash \varphi[\bar{a}]\}$. Let $\Delta \subseteq \Gamma$ be finite and nonempty. Then since $u \in R^{\mathcal{M}}[v]$, $\mathcal{M}, w, v \Vdash \Diamond \bigwedge \Delta[\bar{a}]$. Since by hypothesis $\mathcal{M}, w, v, \bar{a} \equiv_{S_1,\ldots,S_n} \mathcal{N}, w', v', \bar{b}$ (and since for each $\psi \in \Delta$, we could replace $\bar{z}$ with fresh new variables in VAR), it follows that $\mathcal{N}, w', v' \Vdash \Diamond \bigwedge \Delta[\bar{b}]$. Hence, Δ is satisfiable in $\{w'\} \times R^{\mathcal{N}}[v']$ over $\bar{b}$. By $\mathcal{L}^{1M}(S_1, \ldots, S_n)$-saturation, there is a $u' \in R^{\mathcal{N}}[v']$ such that $\mathcal{M}, w, u' \Vdash \Gamma[\bar{b}]$. Thus, $\mathcal{M}, w, u, \bar{a} \equiv_{S_1,\ldots,S_n} \mathcal{N}, w', u', \bar{b}$. Likewise for the Zag clause. ✓

Back-Forth. Let $a' \in \delta^{\mathcal{M}}(v)$. Define $\Gamma(x, \bar{z}) := \{\varphi(x, \bar{z}) \mid \mathcal{M}, w, v \Vdash \varphi[a', \bar{a}]\}$. Let $\Delta \subseteq \Gamma$ be finite and nonempty. Then $\mathcal{M}, w, v \Vdash \exists x \bigwedge \Delta(x, \bar{z})[\bar{a}]$. Since by hypothesis $\mathcal{M}, w, v, \bar{a} \equiv_{S_1,\ldots,S_n} \mathcal{N}, w', v', \bar{b}$, it follows that $\mathcal{N}, w', v' \Vdash \exists x \bigwedge \Delta(x, \bar{z})[\bar{b}]$. Hence, Δ is $\exists$-satisfiable in $\{\langle w', v'\rangle\}$ over $\bar{b}$ with respect to $\mathcal{N}$. By $\mathcal{L}^{1M}(S_1, \ldots, S_n)$-saturation, Γ itself is $\exists$-satisfiable in $\{\langle w', v'\rangle\}$ over $\bar{b}$ with respect to $\mathcal{N}$. So there is a $b' \in \delta^{\mathcal{M}}(v')$ such that $\mathcal{N}, w', v' \Vdash \Gamma[b', \bar{b}]$. Thus, $\mathcal{M}, w, v, \bar{a}, a' \equiv_{S_1,\ldots,S_n} \mathcal{N}, w', v', \bar{b}, b'$. Likewise for the Forth clause. ✓

The $\mathcal{F}$-Zig-Zag-clauses are just like the Zig-Zag clause above, and the other quantifier Back-Forth clauses are just like the Back-Forth clauses above. The Act and Diag clauses are taken care of automatically by the fact that $\mathcal{M}, w, v, \bar{a} \equiv_{S_1,\ldots,S_n} \mathcal{N}, w', v', \bar{b}$ (assuming @/↓ is among $S_1, \ldots, S_n$). ∎

Definition A.4 (*Realization*). Let $\mathcal{L}^{TS}(\bar{z}, \bar{t})$ be $\mathcal{L}^{TS}$ extended with $\bar{z} \notin$ VAR and $\bar{t} \notin$ SVAR. Let $\bar{b} \in D$ and $\bar{v} \in W$ where $|\bar{b}| = |\bar{z}|$ and $|\bar{v}| = |\bar{t}|$. Let $\mathcal{M}$ be a model, and let $\Gamma(\bar{x}, \bar{z}; \bar{s}, \bar{t})$ be a set of $\mathcal{L}^{TS}(\bar{z}, \bar{t})$-formulas whose only free variables are among $\bar{x}, \bar{z}, \bar{s}, \bar{t}$.

- Γ is ***realized over $\bar{b}$ and $\bar{v}$ (with respect to $\mathcal{M}$)*** if there are some $\bar{a} \in D$ and $\bar{u} \in W$ such that $\mathcal{M} \models \Gamma[\bar{a}, \bar{b}; \bar{u}, \bar{v}]$. We call $\bar{b}$ and $\bar{v}$ ***parameters***.

- Γ is ***finitely realized over $\bar{b}$ and $\bar{v}$ (with respect to $\mathcal{M}$)*** if every finite subset of Γ is realized over $\bar{b}$ and $\bar{v}$ (with respect to $\mathcal{M}$).

Definition A.5 (*Saturation*). We will say $\mathcal{M}$ is ***countably saturated*** if for every finite $\bar{b} \in D$, every finite $\bar{v} \in W$, and every set Γ of $\mathcal{L}^{TS}(\bar{z}, \bar{t})$-formulas (where $|\bar{b}| = |\bar{z}|$ and $|\bar{v}| = |\bar{t}|$) that is finitely realized over $\bar{b}$ and $\bar{v}$, Γ is also realized over $\bar{b}$ and $\bar{v}$.

Lemma A.6 *(Countable Saturation Implies Modal Saturation).* If $\mathcal{M}$ is countably saturated, then it is $\mathcal{L}$-saturated.

Proof: Let $\mathcal{M}$ be a countably saturated model. Suppose a set $\Gamma(\overline{x},\overline{z})$ of $\mathcal{L}(\overline{z})$-formulas is finitely $\exists$-satisfiable in $\{w\} \times R[v]$ over $\overline{b}$. Consider the set:

$$\Gamma^*(\overline{x},\overline{z}; s, r, t) := \mathsf{ST}_{s,t}(\Gamma) \cup \{\mathsf{R}(r,t)\} \cup \{\mathsf{E}(x_i; t) \mid x_i \in \overline{x}\}.$$

Let $\Delta \subseteq \Gamma$ be finite and nonempty. Since Δ is $\exists$-satisfiable in $\{w\} \times R[v]$, there is a $u \in R[v]$ and some $\overline{a} \in \delta(u)$ such that $\mathcal{M}, w, u \Vdash \Delta[\overline{a}, \overline{b}]$. Let:

$$\Delta^* := \mathsf{ST}_{s,t}(\Delta) \cup \{\mathsf{R}(r,t)\} \cup \{\mathsf{E}(x_i; t) \mid x_i \in \overline{x}\}.$$

Then $\mathcal{M} \models \Delta^*[\overline{a}, \overline{b}; w, v, u]$. But $\Delta^* \subseteq \Gamma^*$ is finite. So every finite subset of Γ^* is realized over $\overline{b}, w, v$, and u. By countable saturation, there are some $\overline{a} \in D$ and $u \in W$ such that $\mathcal{M} \models \Gamma^*[\overline{a}, \overline{b}; w, v, u]$. But then $u \in R[v]$, $\overline{a} \in \delta(u)$, and $\mathcal{M}, w, u \Vdash \Gamma[\overline{a}, \overline{b}]$. So Γ is $\exists$-satisfiable in $\{w\} \times R[v]$. Likewise for $\exists$-satisfiability in $\{\langle w, v \rangle\}$, except you define:

$$\Gamma^*(\overline{x},\overline{z}; s, t) := \{\mathsf{ST}_{s,t}(\varphi) \mid \varphi \in \Gamma\} \cup \{\mathsf{E}(x_i; t) \mid x_i \in \overline{x}\}.$$

Similarly for the other kinds of satisfiability.   ∎

It follows from **Lemma A.3** and **Lemma A.6** that:

Corollary A.7 *(Countable Saturation implies Hennessy-Miller Property).* If $\mathcal{M}$ and $\mathcal{N}$ are countably saturated, and it $\mathcal{M}, w, v, \overline{a} \equiv_{S_1,\dots,S_n} \mathcal{N}, w', v', \overline{b}$, then it follows that $\mathcal{M}, w, v, \overline{a} \leftrightarrows_{S_1,\dots,S_n} \mathcal{N}, w', v', \overline{b}$.

Definition A.8 *(Ultraproducts).* Let $N \neq \varnothing$. An **ultrafilter over** N is a set $U \subseteq \wp(N)$ where U is closed under supersets and finite intersections, $\varnothing \notin U$, and for all $S \in \wp(N)$, either $S \in U$ or $\overline{S} \in U$. Let U be an ultrafilter over N. For each $i \in N$, let $W_i \neq \varnothing$. Then $\prod_{i \in N} W_i$ is the set of functions $f \colon N \to \bigcup_{i \in N} W_i$ where $f(i) \in W_i$. We will say $f \sim_U f'$ if $\{i \in N \mid f(i) = f'(i)\} \in U$. Define $[f]_U = \{f' \mid f \sim_U f'\}$. Finally, we will define the **ultraproduct of** W_i **modulo** U as the set $\prod_U W_i := \{[f]_U \mid f \in \prod_{i \in N} W_i\}$. We will drop the subscript U when the ultrafilter in question is obvious from the context. An **ultrapower** is an ultraproduct where $W_i = W$ for all $i \in N$, which we may write as $\prod_U W$.

Definition A.9 *(Ultrapowers of Models).* Let $\mathcal{M} = \langle W, R, F, D, \delta, I \rangle$ be a model. The **ultrapower model of** $\mathcal{M}$ **modulo** U is the model $\prod_U \mathcal{M}$ defined as follows:

- $W_U \coloneqq \prod_U W$

- $R_U([f_1],[f_2])$ iff $\{i \in N \mid R(f_1(i), f_2(i))\} \in U$

- $F_U([f_1],[f_2])$ iff $\{i \in N \mid F(f_1(i), f_2(i))\} \in U$

- $D_U \coloneqq \prod_U D$

- $[o] \in \delta_U([f])$ iff $\{i \in N \mid o(i) \in \delta(f(i))\} \in U$

- $\langle [o_1], \dots, [o_n] \rangle \in I_U(P, [f])$ iff $\{i \in N \mid \langle o_1(i), \dots, o_n(i)\rangle \in I(P, f(i))\} \in U$.

It is a routine exercise to show that these definitions are well-defined, i.e., they do not depend on the representative of the equivalence class used in their statement.

> **Theorem A.10** *(Łoś's Theorem).* The following are equivalent:
>
> (a) $\prod_U \mathcal{M} \models \alpha[[o_1], \dots, [o_n]]$.
>
> (b) $\{i \in N \mid \mathcal{M} \models \alpha[o_1(i), \dots, o_n(i)]\} \in U$.

This can be proven by induction.[34] Now, define $f_w \colon i \mapsto w$ and $o_a \colon i \mapsto a$. If g is a variable assignment over $\mathcal{M}$, define $g_U \colon x \mapsto o_{g(x)}$. Let the **diagonal map** be the map d from $\mathcal{M}$ to $\prod_U \mathcal{M}$ such that $d(w) = f_w$ and $d(a) = o_a$. Then it is a straightforward corollary of **Theorem A.10** that the diagonal map is an elementary embedding of $\mathcal{M}$ into $\prod_U \mathcal{M}$.

Say that U is **countably incomplete** if there is a countable subset of U whose intersection is not in U. A standard result from model theory shows that if U is countably incomplete, then $\prod_U \mathcal{M}$ is countably saturated.[35] The important point, however, is that we can always find a countably saturated elementary extension of $\mathcal{M}$ (viz., $\prod_U \mathcal{M}$ where U is a countably incomplete ultrafilter). This is all we will need below.

> **Proof** *(Theorem 3.5)*:   Let $\Gamma \coloneqq \{\mathsf{ST}_{s,t}(\varphi) \mid \alpha \models \mathsf{ST}_{s,t}(\varphi)\}$. It suffices to show (by the compactness of $\mathcal{L}^{\mathsf{TS}}$) that $\Gamma \models \alpha$. Suppose $\mathcal{M}, g \models \Gamma$. Define:
>
> $$\Delta \coloneqq \{\mathsf{ST}_{s,t}(\varphi) \mid \mathcal{M}, g \models \mathsf{ST}_{s,t}(\varphi)\} \cup \{\alpha\}.$$
>
> It is easy to show that Δ is satisfiable (again by compactness). Let $\mathcal{N}, h \models \Delta$. Then $\mathcal{M}, g(s), g(t), g(\overline{x}) \equiv_{S_1, \dots, S_n} \mathcal{N}, h(s), h(t), h(\overline{x})$ by the way we defined Δ. Now, by the results above, there exist elementary extensions $e \colon \mathcal{M} \preccurlyeq \mathcal{M}'$ and $f \colon \mathcal{N} \preccurlyeq \mathcal{M}'$ that are countably saturated. Since these are elementary embeddings:
>
> $$\mathcal{M}', e(g(s)), e(g(t)), e(g(\overline{x})) \equiv_{S_1, \dots, S_n} \mathcal{N}', f(h(s)), f(h(t)), f(h(\overline{x})).$$
>
> By **Corollary A.7**, since these are countably saturated:
>
> $$\mathcal{M}', e(g(s)), e(g(t)), e(g(\overline{x})) \leftrightarrows_{S_1, \dots, S_n} \mathcal{N}', f(h(s)), f(h(t)), f(h(\overline{x})).$$

---

[34]See Chang and Keisler [1990, Chp. 4].
[35]See Bell and Slomson [2006, pp. 222–224].

Hence, by invariance under bisimulation, $\mathcal{M}', g' \models \alpha$ (where $g'(x) = e(g(x))$). Since $e : \mathcal{M} \preccurlyeq \mathcal{M}'$, it follows that $\mathcal{M}, g \models \alpha$.   ∎

§B   Bisimulation Proofs

In this appendix, we give more formal details regarding the bisimulation proofs from §4. We start with the proof of **Proposition 4.2**. Before reading, recall the definition of a partial isomorphism from **Definition 4.1**, and the definition of the models $\mathcal{E}_1$ and $\mathcal{E}_2$ (see Figure 2 for a picture). Note that $\mathcal{M}, w, v, \overline{a} \simeq \mathcal{N}, w', v', \overline{b}$ iff the map $a_i \mapsto b_i$ is a partial isomorphism between them. In particular, for our models $\mathcal{E}_1$ and $\mathcal{E}_2$, if $a_i \in \delta_1(v_1)$ iff $b_i \in \delta_2(v_2)$ and $a_i = a_j$ iff $b_i = b_j$, then $\mathcal{E}_1, w_{\mathbb{N}}, v_1, \overline{a} \simeq \mathcal{E}_2, w_{\mathbb{N}}, v_2, \overline{b}$, since this means $a_i \mapsto b_i$ is a partial isomorphism. We will make use of this implicitly throughout.

We will define our bisimulation in stages. First, set $Z_0 = \{\langle w_{\mathbb{N}}, w_{\mathbb{N}}; w_{\mathbb{N}}, w_{\mathbb{N}}\rangle\}$. Next, suppose we have constructed Z_i so that for all $\langle w_{\mathbb{N}}, v_1, \overline{a}; w_{\mathbb{N}}, v_2, \overline{b}\rangle \in Z_i$, $|\delta_1(v_1)| = |\delta_2(v_2)|$ and $\mathcal{E}_1, w_{\mathbb{N}}, v_1, \overline{a} \simeq \mathcal{E}_2, w_{\mathbb{N}}, v_2, \overline{b}$. (Clearly this holds for Z_0.) Define the following sets:

$$Z_i^{\text{ZZ-fin}} = \left\{ \langle w_{\mathbb{N}}, w_S, \overline{a}; w_{\mathbb{N}}, w_T, \overline{b}\rangle \; \middle| \; \begin{array}{l} \text{(i) } \exists u_1, u_2 \colon \langle w_{\mathbb{N}}, u_1, \overline{a}; w_{\mathbb{N}}, u_2, \overline{b}\rangle \in Z_i \\ \text{(ii) } S \subseteq \mathbb{N} \text{ is finite and nonempty} \\ \text{(iii) } T \subseteq \mathbb{N}^{\infty} \text{ is finite and nonempty} \\ \text{(iv) } a_i \in S \text{ iff } b_i \in T \text{ (v) } |S| = |T| \end{array} \right\}$$

$$Z_i^{\text{ZZ-cofin}} = \left\{ \langle w_{\mathbb{N}}, w_{\mathbb{N}-S}, \overline{a}; w_{\mathbb{N}}, w_{\mathbb{N}^{\infty}-T}, \overline{b}\rangle \; \middle| \; \begin{array}{l} \text{(i) } \exists u_1, u_2 \colon \langle w_{\mathbb{N}}, u_1, \overline{a}; w_{\mathbb{N}}, u_2, \overline{b}\rangle \in Z_i \\ \text{(ii) } S \subseteq \mathbb{N} \text{ is finite (iii) } T \subseteq \mathbb{N}^{\infty} \text{ is finite} \\ \text{(iv) } a_i \in S \text{ iff } b_i \in T \end{array} \right\}$$

$$Z_i^{\text{BF}} = \left\{ \langle w_{\mathbb{N}}, u_1, \overline{a}, a; w_{\mathbb{N}}, u_2, \overline{b}, b\rangle \; \middle| \; \begin{array}{l} \text{(i) } \langle w_{\mathbb{N}}, u_1, \overline{a}; w_{\mathbb{N}}, u_2, \overline{b}\rangle \in Z_i \\ \text{(ii) } a \in \delta_1(u_1) \text{ and } b \in \delta_2(u_2) \\ \text{(iii) } a = a_i \text{ iff } b = b_i \end{array} \right\}.$$

Set $Z_{i+1} \coloneqq Z_i \cup Z_i^{\text{ZZ-fin}} \cup Z_i^{\text{ZZ-cofin}} \cup Z_i^{\text{BF}}$.

Lemma B.1 (*This Construction Can Continue*). If $\langle w_{\mathbb{N}}, v_1, \overline{a}; w_{\mathbb{N}}, v_2, \overline{b}\rangle \in Z_{i+1}$, then we have $|\delta_1(v_1)| = |\delta_2(v_2)|$ and $\mathcal{E}_1, w_{\mathbb{N}}, v_1, \overline{a} \simeq \mathcal{E}_2, w_{\mathbb{N}}, v_2, \overline{b}$.

Proof: Suppose $\langle w_{\mathbb{N}}, v_1, \overline{a}; w_{\mathbb{N}}, v_2, \overline{b}\rangle \in Z_{i+1}$. If $\langle w_{\mathbb{N}}, v_1, \overline{a}; w_{\mathbb{N}}, v_2, \overline{b}\rangle \in Z_i$, then this is obvious. Otherwise, $\langle w_{\mathbb{N}}, v_1, \overline{a}; w_{\mathbb{N}}, v_2, \overline{b}\rangle$ is either in $Z_i^{\text{ZZ-fin}}$, $Z_i^{\text{ZZ-cofin}}$, or Z_i^{BF}.

Case 1: $Z_i^{\text{ZZ-fin}}$. So $v_1 = w_S$ and $v_2 = w_T$ for some S and T. By the fact that $|S| = |T|$, $|\delta_1(v_1)| = |\delta_2(v_2)|$. And by construction, $a_i \in S$ iff $b_i \in T$, so $a_i \in \delta_1(v_1)$ iff $b_i \in \delta_2(v_2)$, and thus $\mathcal{E}_1, w_{\mathbb{N}}, v_1, \overline{a} \simeq \mathcal{E}_2, w_{\mathbb{N}}, v_2, \overline{b}$. ✓

Case 2: $Z_i^{\text{ZZ-cofin}}$. Same reasoning, only we know $|\delta_1(v_1)| = |\delta_2(v_2)| = \aleph_0$. ✓

Case 3: Z_i^{BF}. We already know $|\delta_1(v_1)| = |\delta_2(v_2)|$ since this was guaranteed by Z_i. Moreover, both $a \in \delta_1(v_1)$ and $b \in \delta_2(v_2)$, so we still meet (Existence). And by the fact that $a = a_i$ iff $b = b_i$, we still meet (Equality). ✓   ∎

Lemma B.1 guarantees we can continue the construction. Finally, define $Z := \bigcup_{i \in \omega} Z_i$.

Proof (*Proposition 4.2*): Suppose $\langle w_{\mathbb{N}}, v_1, \overline{a}; w_{\mathbb{N}}, v_2, \overline{b} \rangle \in Z$. Then there is some i such that $\langle w_{\mathbb{N}}, v_1, \overline{a}; w_{\mathbb{N}}, v_2, \overline{b} \rangle \in Z_i$. By **Lemma B.1**, $\mathcal{E}_1, w_{\mathbb{N}}, v_1, \overline{a} \simeq \mathcal{E}_2, w_{\mathbb{N}}, v_2, \overline{b}$, so (Atomic) and (Eq) (as well as (Ex)) are satisfied.

Zig. Let $u_1 \in W_1$. Suppose that $u_1 = w_S$ for some finite nonempty $S \subseteq \mathbb{N}$. We want to show that there is a $T \subseteq \mathbb{N}^\infty$ such that $\langle w_{\mathbb{N}}, w_S, \overline{a}; w_{\mathbb{N}}, w_T, \overline{b} \rangle \in Z_i^{\mathrm{ZZ\text{-}fin}}$. List the elements $c_1, \ldots, c_n \in S - \{\overline{a}\}$. Let $d_1, \ldots, d_n$ be n-many distinct elements from $\mathbb{N}^\infty - \{\overline{b}\}$, and set $T = \{b_i \mid a_i \in S\} \cup \{d_1, \ldots, d_n\}$. Then $T \subseteq \mathbb{N}^\infty$ is also finite and nonempty, $|S| = |T|$, and $a_i \in S$ iff $b_i \in T$. So $\langle w_{\mathbb{N}}, w_S, \overline{a}; w_{\mathbb{N}}, w_T, \overline{b} \rangle \in Z_i^{\mathrm{ZZ\text{-}fin}}$. The case where $u_1 = w_{\mathbb{N}-S}$ is essentially the same, except one does not need $|S| = |T|$. ✓

Zag. As above. ✓

Forth. Let $a \in \delta_1(v_1)$. If $a = a_i$, then we can just pick $b = b_i$, and just note that $\langle w_{\mathbb{N}}, v_1, \overline{a}, a_i; w_{\mathbb{N}}, v_2, \overline{b}, b_i \rangle \in Z_i^{\mathrm{BF}}$. So suppose a is not among $\overline{a}$. Since $|\delta_1(v_1)| = |\delta_2(v_2)|$, we have that $|\delta_1(v_1) - \{\overline{a}\}| = |\delta_2(v_2) - \{\overline{b}\}|$. And the former is not empty since $a \in \delta_1(v_1) - \{\overline{a}\}$. So pick any $b \in \delta_2(v_2) - \{\overline{b}\}$. Then we have that $\langle w_{\mathbb{N}}, v_1, \overline{a}, a; w_{\mathbb{N}}, v_2, \overline{b}, b \rangle \in Z_i^{\mathrm{BF}}$. ✓

Back. As above. ✓   ∎

Now for the proof of **Proposition 4.3**. As before, set $Z_0 = \{\langle w, w; w, w \rangle\}$. Now suppose we have constructed Z_i and for all $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \in Z_i$, $\mathcal{R}_1, w, u_1, \overline{a} \simeq \mathcal{R}_2, w, \overline{u}_2, \overline{b}$ and $u_1 = w$ iff $u_2 = w$. Define the following sets:

$$Z_i^{\mathrm{Act}} = \{\langle w, w, \overline{a}; w, w, \overline{b} \rangle \mid \exists u_1, u_2 \colon \langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \in Z_i \}$$

$$Z_i^{\mathrm{ZZ}} = \left\{ \langle w, v_S, \overline{a}; w, v_T, \overline{b} \rangle \;\middle|\; \begin{array}{l} \text{(i)} \;\; \exists u_1, u_2 \colon \;\; \langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \;\; \in \;\; Z_i \\ \text{(ii)} \;\; S, T \;\; \subseteq \;\; \mathbb{N} \;\; \text{are finite and nonempty} \\ \text{(iii)} \;\; a_i \in S \text{ iff } b_i \in T \end{array} \right\}$$

$$Z_i^{\mathrm{ZZ\varnothing}} = \left\{ \langle w, v_S, \overline{a}; w, v_\varnothing, \overline{b} \rangle \;\middle|\; \begin{array}{l} \text{(i)} \;\; \exists u_1, u_2 \colon \;\; \langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \;\; \in \;\; Z_i \\ \text{(ii)} \;\; S \;\; \subseteq \;\; \mathbb{N} \;\; \text{is finite and nonempty} \\ \text{(iii)} \;\; S \cap \{\overline{a}\} = \varnothing \end{array} \right\}$$

$$Z_i^{\mathrm{BF}} = \left\{ \langle w, u_1, \overline{a}, a; w, u_2, \overline{b}, b \rangle \;\middle|\; \begin{array}{l} \text{(i)} \;\; \langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \;\; \in \;\; Z_i \\ \text{(ii)} \;\; a \in \delta_1(u_1) \text{ and } b \in \delta_2(u_2) \\ \text{(iii)} \;\; a = a_i \text{ iff } b = b_i \quad \text{(iv)} \; a \in \mathbb{N} \\ \text{iff } b \in \mathbb{N} \end{array} \right\}.$$

Then set: $Z_{i+1} = Z_i \cup Z_i^{\mathrm{Act}} \cup Z_i^{\mathrm{ZZ}} \cup Z_i^{\mathrm{ZZ\varnothing}} \cup Z_i^{\mathrm{BF}}$.

Lemma B.2 (*This Construction Can Continue Too*). If $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \in Z_{i+1}$, then $\mathcal{R}_1, w, u_1, \overline{a} \simeq \mathcal{R}_2, w, \overline{u}_2, \overline{b}$ and $u_1 = w$ iff $u_2 = w$.

Proof: Suppose $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \in Z_{i+1}$. It is easy to verify that $u_1 = w$ iff $u_2 = w$ by looking at the constructions above. If $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \in Z_i$, then we are done. So suppose $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \notin Z_i$.

First, (Predicate). If $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \in Z_i^{\mathrm{Act}} \cup Z_i^{\mathrm{ZZ}} \cup Z_i^{\mathrm{ZZ}\varnothing}$, then we know that $\exists u_1', u_2'\colon \ \langle w, u_1', \overline{a}; w, u_2', \overline{b} \rangle \in Z_i$. So $\mathcal{R}_1, w, u_1', \overline{a} \simeq \mathcal{R}_2, w, u_2', \overline{b}$. But since $u_1' = w$ iff $u_2' = w$, that means $a_i \in \mathbb{N}$ iff $b_i \in \mathbb{N}$. So since $u_1 = w$ iff $u_2 = w$, $\mathcal{R}_1, w, u_1, \overline{a}$ and $\mathcal{R}_2, w, u_2, \overline{b}$ satisfy (Predicate). If instead $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle = \langle w, u_1, \overline{c}, c; w, u_2, \overline{d}, d \rangle \in Z_i^{\mathrm{BF}}$, then $\langle w, u_1, \overline{c}; w, u_2, \overline{d} \rangle \in Z_i$, which (by the same reasoning as above) means $c_i \in \mathbb{N}$ iff $d_i \in \mathbb{N}$. And by construction of Z_i^{BF}, $c \in \mathbb{N}$ iff $d \in \mathbb{N}$. So since $u_1 = w$ iff $u_2 = w$, again $\mathcal{R}_1, w, u_1, \overline{a}$ and $\mathcal{R}_2, w, u_2, \overline{b}$ satisfy (Predicate).

Next, (Existence). This is trivial if $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \in Z_i^{\mathrm{Act}}$, since $\delta_1(w) = \delta_2(w) = \mathbb{Z}$. It is guaranteed by construction in all other cases.

Finally, (Equality). This is trivial in every case, except Z_i^{BF}, in which case it is guaranteed by construction.   ∎

As before, define $Z := \bigcup_{i \in \omega} Z_i$.

Proof (*Proposition 4.3*): Suppose $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \in Z$. Then there is an i such that $\langle w, u_1, \overline{a}; w, u_2, \overline{b} \rangle \in Z_i$. By **Lemma B.2**, (Atomic) and (Eq) (as well as (Ex)) are satisfied.

Act. By construction of Z_i^{Act}, $\langle w, w, \overline{a}; w, w, \overline{b} \rangle \in Z_{i+1}$. ✓

Zig. Let $u_1' \in W_1$. If $u_1' = w$, then this is covered by the above case. So let $u_1' = v_S$ instead. Define $T := \{ b_i \mid a_i \in S \} \cup \{c\}$ where $c \in \mathbb{N} - \{\overline{b}\}$ is arbitrary. Then T is finite and nonempty, and $a_i \in S$ iff $b_i \in T$. So $\langle w, v_S, \overline{a}; w, v_T, \overline{b} \rangle \in Z_i^{\mathrm{ZZ}}$. ✓

Zag. As above, except in the case where we pick $v_\varnothing \in W_2$. In that case, let $S \subseteq \mathbb{N}$ be any finite nonempty set such that $S \cap \{\overline{a}\} = \varnothing$. Then $\langle w, v_S, \overline{a}; w, v_\varnothing, \overline{b} \rangle \in Z_i^{\mathrm{ZZ}\varnothing}$. ✓

Back. Let $a \in \delta_1(u_1)$ (we assume without loss of generality that $a \notin \{\overline{a}\}$). If $a \in \mathbb{N}^-$, then pick any new $b \in \mathbb{N}^-$. If instead $a \in \mathbb{N}$, then since $\delta_2(u_2) \cap \mathbb{N}$ is infinite, pick any new $b \in \delta_2(u_2) \cap \mathbb{N}$. Either way, $\langle w, u_1, \overline{a}, a; w, u_2, \overline{b}, b \rangle \in Z_i^{\mathrm{BF}}$. ✓

Forth. As above. ✓   ∎

Other inexpressibility proofs are straightforward once the winning strategy for Eloïse is worked out.

§C   Inexpressibility in $\mathcal{L}^{1M}(\approx, @, \downarrow, \mathcal{F})$

In this appendix, we will prove that (4) is not expressible as an $\mathcal{L}^{1M}(\approx, @, \downarrow, \mathcal{F})$-formula. Recall (4):

$$\exists t \ (\mathsf{R}(s^*, t) \wedge \forall x \ (\mathsf{Rich}(x; s^*) \rightarrow \mathsf{Poor}(x; t))). \tag{4}$$

First, we define our models $\mathcal{R}_3 = \langle W_3, R_3, F_3, D_3, \delta_3, I_3 \rangle$ and $\mathcal{R}_4 = \langle W_4, R_4, F_4, D_4, \delta_4, I_4 \rangle$. Our global domains will contain $\mathbb{Z}$ plus a disjoint copy of $\mathbb{N}$, which we will call $\mathbb{N}_\infty :=$ $\{\infty_n \mid n \in \mathbb{N}\}$. So $D_3 = D_4 = \mathbb{Z} \cup \mathbb{N}_\infty$. All the accessibility relations are universal in their respective models. If $T \subseteq \mathbb{N}$, let $T_\infty := \{\infty_n \mid n \in T\}$. **Please note**: throughout this section, when we write $\overline{T_\infty}$, we mean $\mathbb{N}_\infty - T$; when we write $\overline{S}$ where $S \subseteq \mathbb{N}$, we mean $\mathbb{N} - S$.

For each finite *nonempty* $S \subseteq \mathbb{N}$, and each finite $T \subseteq \mathbb{N}$, W_1 will contain a world w_S^T and a world v_S^T. Intuitively, w_S^T is a world where (i) every negative integer is poor, (ii) every integer of $(\mathbb{N} - S)$ is rich, (iii) every object of T_∞ is rich, and (iv) nothing in $S \cup \overline{T_\infty}$ exists. v_S^T is like w_S^T except the rich and the poor are flipped. In addition, for any finite $T \subseteq \mathbb{N}$, there will be a world of the form $w_\varnothing^T$ in W_1 (our actual world will be $w := w_\varnothing^\varnothing$). W_2 is like W_1 except it also contains worlds of the form $v_\varnothing^T$. See Figure 7 for a picture.

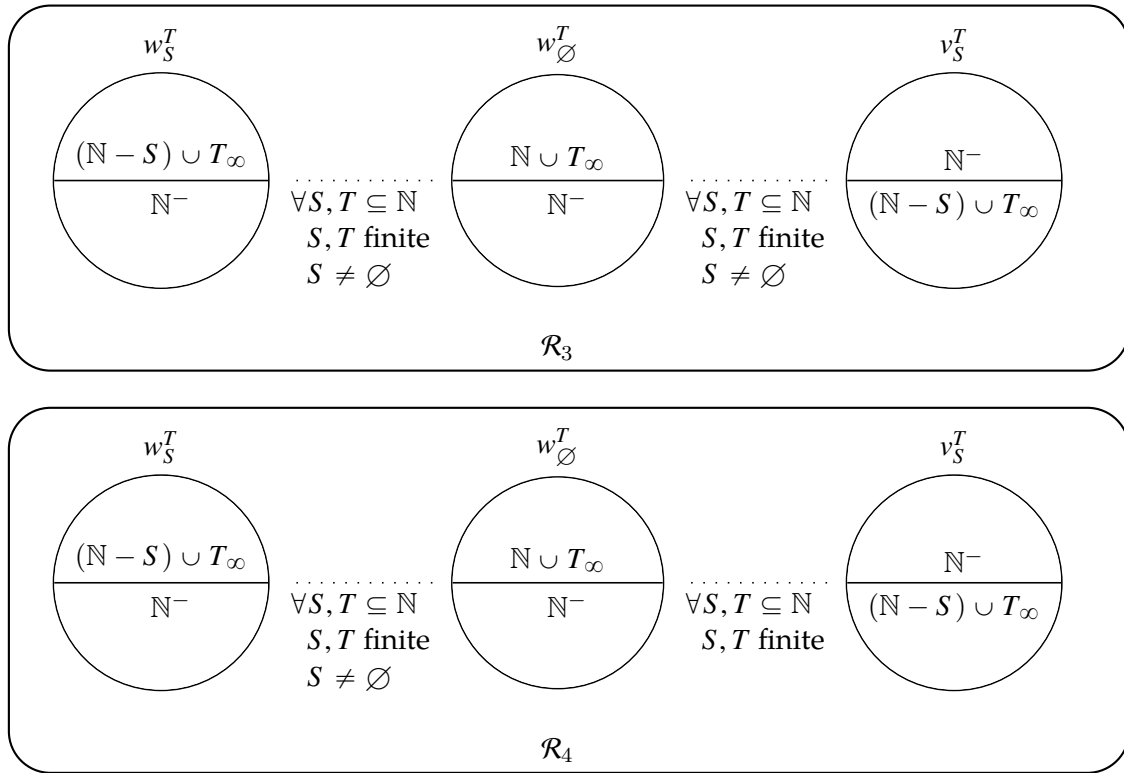Figure 7: $\mathcal{L}^{1M}(\approx, @, \downarrow, \mathcal{F})$-bisimilar models that still disagree on (R).

Observe $\mathcal{R}_3 \not\models (4)[w]$ while $\mathcal{R}_4 \models (4)[w]$. Furthermore, recall that the reason $\mathcal{R}_1$ and $\mathcal{R}_2$ could not be used to show that (4) is not expressible as an $\mathcal{L}^{1M}(\approx, @, \downarrow, \mathcal{F})$-formula was because they disagreed on the following formulas at w:

$$\exists x \, (\mathsf{Rich}(x) \wedge \Diamond \downarrow (\mathsf{Poor}(x) \wedge \Box \forall y \, @\mathsf{E}(y))) \tag{8}$$

$$\exists x \, (\mathsf{Rich}(x) \wedge \langle \mathcal{F} \rangle @ \, (\mathsf{Poor}(x) \wedge \Box \forall y \, @\mathsf{E}(y))) . \tag{9}$$

Observe that this is no longer the case: w does not satisfy either (8) or (9) in $\mathcal{R}_3$ or $\mathcal{R}_4$.

Proposition C.1 (*Strengthened Inexpressibility of (R)*). $\mathcal{R}_3, w, w \leftrightarrows_{\approx, @, \downarrow, \mathcal{F}} \mathcal{R}_4, w, w$.

Proof: Clearly $\mathcal{R}_3, w, w \simeq \mathcal{R}_4, w, w$. So suppose that $\mathcal{R}_3, s_{S_1}^{T_1}, t_{S_2}^{T_2}, \bar{a} \simeq \mathcal{R}_4, s_{S_1'}^{T_1'}, t_{S_2'}^{T_2'}, \bar{b}$, where:

(I) $s_{S_1}^{T_1} = w_{S_1}^{T_1}$ iff $s_{S_1'}^{T_1'} = w_{S_1'}^{T_1'}$ (and likewise for t)

(II) $a_i \in \mathbb{N}^-$ iff $b_i \in \mathbb{N}^-$

(III) $a_i \in S_1 \cup \overline{T_{1\infty}}$ iff $b_i \in S_1' \cup \overline{T_{1\infty}'}$

(IV) $a_i \in S_2 \cup \overline{T_{2\infty}}$ iff $b_i \in S_2' \cup \overline{T_{2\infty}'}$

(V) $|(S_1 - S_2) \cup (T_{2\infty} - T_{1\infty})| = |(S_1' - S_2') \cup (T_{2\infty}' - T_{1\infty}')|$.

Notice in particular that $\mathcal{R}_3, w, w \simeq \mathcal{R}_4, w, w$ meets all of these constraints vacuously. We will show using (I)–(V) that, regardless of Abelard's move, Eloïse can continue the game in a way that preserves (I)–(V). Note throughout that if I use the same letter, say u, for u_S^T and $u_{S'}^{T'}$, I mean for u_S^T to be a w-world iff $u_{S'}^{T'}$ is a w-world.

First, suppose Abelard decides to pick an object $a \in \delta_3(t_{S_2}^{T_2})$ (the case where he picks a $b \in \delta_4(t_{S_2'}^{T_2'})$ is symmetric). If he does this, then obviously (I) and (V) are met regardless of what Eloïse picks. So she just needs to ensure (II)–(IV) are met. If $a \in \mathbb{N}^-$, then Eloïse can pick an arbitrary $b \in \mathbb{N}^-$ that has not already been picked. Otherwise, since $a \in \delta_3(t_{S_2}^{T_2})$, that means $a \notin S_2 \cup \overline{T_{2\infty}}$. There are two cases to consider:

Case 1: $a \notin S_1 \cup \overline{T_{1\infty}}$. That means $a \in (\mathbb{N} - (S_1 \cup S_2)) \cup (T_{1\infty} \cup T_{2\infty})$. But since S_1', S_2', T_1', and T_2' are all finite, there will be infinitely many $b \in (\mathbb{N} - (S_1' \cup S_2')) \cup (T_{1\infty}' \cup T_{2\infty}')$ that have not been picked yet. So Eloïse can just pick an arbitrary one of those, in which case $b \notin S_1' \cup \overline{T_{1\infty}'}$ and $b \notin S_2' \cup \overline{T_{2\infty}'}$. ✓

Case 2: $a \in S_1 \cup \overline{T_{1\infty}}$. That means we need to ensure that $b \in S_1' \cup \overline{T_{1\infty}'}$ while also ensuring that $b \notin S_2' \cup \overline{T_{2\infty}'}$. That means we need:

$$b \in (S_1' \cup \overline{T_{1\infty}'}) - (S_2' \cup \overline{T_{2\infty}'})$$
$$= (S_1' - S_2') \cup (\overline{T_{1\infty}'} - \overline{T_{2\infty}'})$$
$$= (S_1' - S_2') \cup (T_{2\infty}' - T_{1\infty}').$$

But since $|(S_1 - S_2) \cup (T_{2\infty} - T_{1\infty})| = |(S_1' - S_2') \cup (T_{2\infty}' - T_{1\infty}')|$, and since (II)–(IV) hold for $\bar{a}$ and $\bar{b}$, it is easy to show that:

$$\left|[(S_1 - S_2) \cup (T_{2\infty} - T_{1\infty})] - \{\bar{a}\}\right| = \left|[(S_1' - S_2') \cup (T_{2\infty}' - T_{1\infty}')] - \{\bar{b}\}\right|$$

Hence, there must be some $b \in (S_1' - S_2') \cup (T_{2\infty}' - T_{1\infty}')$ that has not been picked yet. So Eloïse can just pick an arbitrary one of those. ✓

Next, suppose Abelard decides to relocate the game. If he uses the @ or $\downarrow$ moves, then the constraints will all be vacuously satisfied. So suppose he decides to relocate the game in $\mathcal{R}_3$ to $\langle s_{S_1}^{T_1}, u_{S_3}^{T_3} \rangle$. If $T_3 = T_1$ and $S_3 = S_1$, then Eloïse should pick $u_{S_1'}^{T_1'}$. If $T_3 = T_2$ and $S_3 = S_2$, then she should pick $u_{S_2'}^{T_2'}$. Otherwise, Eloïse can pick a T_3' and S_3' using a different method as follows. Define the following sets:

$$S_3^* := \{b_i \in \overline{b} \mid b_i \in \mathbb{N} \text{ and } a_i \in S_3 \cup \overline{T_{1\infty}}\}$$
$$T_{3\infty}^* := \{b_i \in \overline{b} \mid b_i \in \mathbb{N}_\infty \text{ and } a_i \notin S_3 \cup \overline{T_{1\infty}}\}.$$

One can verify that if $S_3' \subseteq \mathbb{N}$ and $T_{3\infty}' \subseteq \mathbb{N}_\infty$ such that $S_3' \cap \{\overline{b}\} = S_3^*$ and $T_{3\infty}' \cap \{\overline{b}\} = T_{3\infty}^*$, then $a_i \in S_3 \cup \overline{T_{3\infty}}$ iff $b_i \in S_3' \cup \overline{T_{3\infty}'}$. We will now show the following:

> **Claim**: There are $S_3' \subseteq \mathbb{N}$ and $T_{3\infty}' \subseteq \mathbb{N}_\infty$ such that:
>
> (i) $S_3' \cap \{\overline{b}\} = S_3^*$
>
> (ii) $T_{3\infty}' \cap \{\overline{b}\} = T_{3\infty}^*$, and
>
> (iii) $|(S_1 - S_3) \cup (T_{3\infty} - T_{1\infty})| = |(S_1' - S_3') \cup (T_{3\infty}' - T_{1\infty}')|$.

Suppose not. That is, suppose that every $S_3' \subseteq \mathbb{N}$ and $T_{3\infty}' \subseteq \mathbb{N}_\infty$ that satisfy (i) and (ii) fail to satisfy (iii). We will show that from this assumption, we can derive a contradiction.

First, suppose there is an $S_3' \subseteq \mathbb{N}$ and $T_{3\infty}' \subseteq \mathbb{N}_\infty$ satisfying (i) and (ii) such that $|(S_1 - S_3) \cup (T_{3\infty} - T_{1\infty})| > |(S_1' - S_3') \cup (T_{3\infty}' - T_{1\infty}')|$. Let $n < \omega$ be such that $|(S_1 - S_3) \cup (T_{3\infty} - T_{1\infty})| = |(S_1' - S_3') \cup (T_{3\infty}' - T_{1\infty}')| + n$ (both sets are finite after all). Since $T_{1\infty}'$ and $T_{3\infty}'$ are finite, we can pick n arbitrary objects $\overline{c} \in \mathbb{N}_\infty - (T_{1\infty}' \cup T_{3\infty}' \cup \{\overline{b}\})$ and set $T_{3\infty}'' := T_{3\infty}' \cup \{\overline{c}\}$. But then $|(S_1 - S_3) \cup (T_{3\infty} - T_{1\infty})| = |(S_1' - S_3') \cup (T_{3\infty}'' - T_{1\infty}')|$, and (ii) is still met replacing $T_{3\infty}'$ with $T_{3\infty}''$. ⨳

Hence, it must be that for every $S_3' \subseteq \mathbb{N}$ and $T_{3\infty}' \subseteq \mathbb{N}_\infty$ satisfying (i) and (ii), $|(S_1 - S_3) \cup (T_{3\infty} - T_{1\infty})| < |(S_1' - S_3') \cup (T_{3\infty}' - T_{1\infty}')|$. Now, we can assume without loss of generality that $(S_1' - S_3') \subseteq \{\overline{b}\}$ and $(T_{3\infty}' - T_{1\infty}') \subseteq \{\overline{b}\}$. Here is why. Suppose $(S_1' - S_3') - \{\overline{b}\} \neq \varnothing$. Then pick a $c \in (S_1' - S_3') - \{\overline{b}\}$ and set $S_3'' := S_3' \cup \{c\}$. Then $|S_1' - S_3''| < |S_1' - S_3'|$, so $|(S_1' - S_3'') \cup (T_{3\infty}' - T_{1\infty}')| < |(S_1' - S_3') \cup (T_{3\infty}' - T_{1\infty}')|$. S_3'' still satisfies (i), so by hypothesis, it still must be that $|(S_1 - S_3) \cup (T_{3\infty} - T_{1\infty})| < |(S_1' - S_3'') \cup (T_{3\infty}' - T_{1\infty}')|$. So we can just keep adding objects from $(S_1' - S_3') - \{\overline{b}\}$ to S_3' in this way until $(S_1' - S_3') - \{\overline{b}\} = \varnothing$. Likewise, we can keep removing objects in $T_{3\infty}'$ from $(T_{3\infty}' - T_{1\infty}') - \{\overline{b}\}$ until $(T_{3\infty}' - T_{1\infty}') - \{\overline{b}\} = \varnothing$.

Thus, we may assume that $(S_1' - S_3') \subseteq \{\overline{b}\}$ and $(T_{3\infty}' - T_{1\infty}') \subseteq \{\overline{b}\}$. It follows that $(S_1' - S_3') \cup (T_{3\infty}' - T_{1\infty}') \subseteq \{\overline{b}\}$. But if $b_i \in (S_1' - S_3') \cup (T_{3\infty}' - T_{1\infty}')$, then $a_i \in (S_1 - S_3) \cup (T_{3\infty} - T_{1\infty})$ by (III) and by the fact that (i) and (ii) imply that $a_i \in S_3 \cup \overline{T_{3\infty}}$ iff $b_i \in S_3' \cup \overline{T_{3\infty}'}$. This gives rise to an injection from $(S_1' - S_3') \cup (\overline{T_{1\infty}'} \cap T_{3\infty}')$ to $(S_1 - S_3) \cup (\overline{T_{1\infty}} \cap T_{3\infty})$, which means $|(S_1 - S_3) \cup (\overline{T_{1\infty}} \cap T_{3\infty})| \geq |(S_1' - S_3') \cup (\overline{T_{1\infty}'} \cap T_{3\infty}')|$. ⨳ This completes our proof of the claim above.

> Thus, Eloïse can just pick any such S'_3 and $T'_{3\infty}$, and it will have the desired properties. If instead Abelard decides to relocate the game in $\mathcal{R}_4$, the strategy is the same: the reasoning above did not rely on Abelard's S_3 being nonempty. Finally, the case where Abelard decides to relocate the game in $\mathcal{R}_3$ to $\langle u^{T_3}_{S_3}, t^{T_2}_{S_1} \rangle$ is symmetric. ∎

§D   MAPPING THE EXPRESSIVE HIERARCHY

In this appendix, we map out in more detail the relative expressive power for various fragments of $\mathcal{L}^{1M}(\approx, @, \downarrow, \mathcal{F}, \Pi)$. We will start by showing that, ignoring E, $\approx$, and Π, the relative expressive power for the remaining languages is accurately diagrammed by Figure 8. This includes the strict inclusions and incomparabilities the diagram suggests. Note that for any class of models $\mathbf{C}$, $\leqslant_\mathbf{C}$ is a preorder.
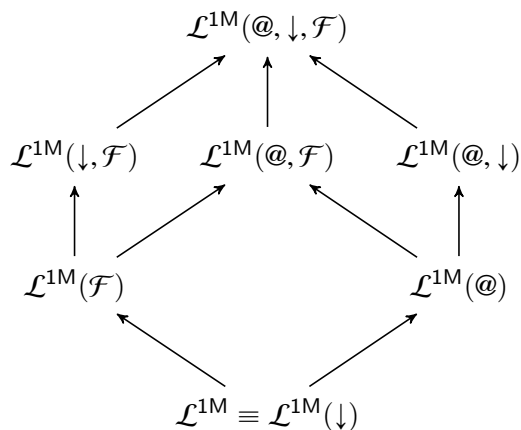

Figure 8: Relative (**D**-)expressive power for languages between $\mathcal{L}^{1M}$ and $\mathcal{L}^{1M}(@, \downarrow, \mathcal{F})$.

> **Lemma D.1** (*Adding Only ↓ Does Nothing*). $\mathcal{L}^{1M} \equiv \mathcal{L}^{1M}(\downarrow)$.

> **Proof:**   First, note that, by induction, for any $\mathcal{L}^{1M}(\downarrow)$-formula φ, $\mathcal{M}, w, v, g \Vdash \varphi$ iff $\mathcal{M}, w', v, g \Vdash \varphi$. Thus, where φ is an $\mathcal{L}^{1M}(\downarrow)$-formula, let φ^- be the result of removing every instance of $\downarrow$ from φ. Then it is easy to show by induction (using this fact for the $\downarrow$-case) that $\Vdash \varphi \leftrightarrow \varphi^-$. Hence, $\mathcal{L}^{1M}(\downarrow) \leqslant \mathcal{L}^{1M}$, and therefore $\mathcal{L}^{1M} \equiv \mathcal{L}^{1M}(\downarrow)$. ∎

Throughout, when I say "$I^\mathcal{M}$ is empty" or "$I^\mathcal{M} = \varnothing$", what I mean is that for all $w \in W^\mathcal{M}$ and all predicates P, $I^\mathcal{M}(P, w) = \varnothing$. Also, if $\bar{a}$ is clear from context, I will use "a_i" to stand for an arbitrary element of $\bar{a}$.

Lemma D.2 (*Adding $\mathcal{F}$*). $\mathcal{L}^{1M}(\mathcal{F}) \nleq_{\mathbf{D}} \mathcal{L}^{1M}(\approx, @, \downarrow, \Pi)$.

Proof: Let $\mathcal{M}_1 = \langle W_1, R_1, F_1, D_1, \delta_1, I_1 \rangle$, where $W_1 = \{w\}$, $D_1 = \delta_1(w) = \{a\}$, $R_1 = F_1 = \{\langle w, w \rangle\}$, and $I_1 = \varnothing$. Let $\mathcal{M}_2$ be just like $\mathcal{M}_1$ except $F_2 = \varnothing$. Then $\mathcal{M}_1, w, w \leftrightarroweq_{\approx, @, \downarrow, \Pi} \mathcal{M}_2, w, w$, but $\mathcal{M}_1, w, w \nVdash \mathcal{F}\bot$ while $\mathcal{M}_2, w, w \Vdash \mathcal{F}\bot$.   ∎

Where $\mathcal{M}$ is a model, let $\mathcal{M}^{\mathsf{E}\to P}$ be the model just like $\mathcal{M}$ except $I^{\mathcal{M}^{\mathsf{E}\to P}}(P, w) = \delta^{\mathcal{M}}(w)$ for all $w \in W^{\mathcal{M}}$. That is, $\mathcal{M}^{\mathsf{E}\to P}$ effectively makes P an existence predicate. Define $\mathcal{M}^{\approx\to P}$ likewise. It will be useful to note the following:

Lemma D.3 (*Replacing E*). If $\mathcal{M}, w, v, \overline{a} \leftrightarroweq_{\mathcal{L}(\mathsf{E})} \mathcal{N}, w', v', \overline{b}$, then $\mathcal{M}^{\mathsf{E}\to P}, w, v, \overline{a} \leftrightarroweq_{\mathcal{L}(\mathsf{E})}$ $\mathcal{N}^{\mathsf{E}\to P}, w', v', \overline{b}$. In addition, if $I^{\mathcal{M}}(P, u) = \varnothing = I^{\mathcal{N}}(P, u')$ for all $u \in W^{\mathcal{M}}$ and all $u' \in W^{\mathcal{N}}$, then the converse holds as well. Likewise for $\approx$ in place of E.

We can use this trick to bootstrap off of previous inexpressibility results which used E or $\approx$ for languages without E or $\approx$. For instance, it is relatively easy to show $\mathcal{L}^{1M}(@) \nleq_{\mathbf{UD}}$ $\mathcal{L}^{1M}(\approx, \downarrow, \mathcal{F})$. Take models $\mathcal{E}_1$ and $\mathcal{E}_2$ from Figure 2. Since $\mathcal{E}_1, w_{\mathbb{N}}, w_{\mathbb{N}} \leftrightarroweq_{\approx} \mathcal{E}_2, w_{\mathbb{N}}, w_{\mathbb{N}}$, we can use **Lemma D.3** to conclude $\mathcal{E}_1^{\mathsf{E}\to P}, w_{\mathbb{N}}, w_{\mathbb{N}} \leftrightarroweq_{\approx} \mathcal{E}_2^{\mathsf{E}\to P}, w_{\mathbb{N}}, w_{\mathbb{N}}$, though they disagree on $\Diamond \exists x \, @ \neg P(x)$.

However, this proof does not show that $\mathcal{L}^{1M}(@) \nleq_{\mathbf{UD}} \mathcal{L}^{1M}(\approx, \downarrow, \mathcal{F}, \Pi)$, since the models are distinguishable by $\Sigma x \neg P(x)$. **Lemma D.4** strengthens this result to include Π, again using **Lemma D.3**:[36]

Lemma D.4 (*Adding @*). $\mathcal{L}^{1M}(@) \nleq_{\mathbf{UD}} \mathcal{L}^{1M}(\approx, \downarrow, \mathcal{F}, \Pi)$.

---

[36]Hodes [1984b, pp. 445–446] claimed to have a proof that $\mathcal{L}^{1M}(\approx, @) \nleq_{\mathbf{UD}} \mathcal{L}^{1M}(\approx, \Pi)$. He also constructs two models which he claims satisfy the same $\mathcal{L}^{1M}(\approx, \Pi)$-formulas, but disagree on the $\mathcal{L}^{1M}(\approx, @)$-formula $\theta_2 = \Box \, (\exists x \, @ \neg \mathsf{E}(x) \to \exists_{\geqslant 2} x \, @ \neg \mathsf{E}(x))$. Here are the models $\mathcal{A}$ and $\mathcal{B}$ he describes (p. 445, his notation; $A(w)$ and $B(w)$ in Hodes's notation means $\delta^{\mathcal{A}}(w)$ and $\delta^{\mathcal{B}}(w)$ in ours, and $\langle w, a \rangle \in V(P)$ in his notation means $a \in I(P, w)$ in ours; he also write $\mathcal{A}, 0 \vDash \varphi$ in place of our $\mathcal{A}, 0, 0 \Vdash \varphi$):

Let $W = \{0\} \cup \{(n, m) \mid n \neq m, n, m \in \omega\}$, $W' = W \cup \{1\}$. Let $A(0) = B(0) = \omega$, $A((n, m)) = B((n, m)) = (\omega - \{n, m\}) \cup \{-n \mid n \in \omega\}$, let $V(P)$ be empty for all $P \in \mathrm{Pred}$, $\mathcal{A} = (W, A, V)$, $\mathcal{B} = (W', B, V)$. Clearly $\mathcal{A}, 0 \vDash \theta_2$ but $\mathcal{B}, 0 \nvDash \theta_2$.

However, Hodes's description of these models is incomplete, since crucially the local domain of 1 is in $\mathcal{B}$ is never specified, and the proof that follows gives no indication of what it might be. Moreover, given the proof requires that $\mathcal{A}, 0, 0 \Vdash \theta_2$ and $\mathcal{B}, 0, 0 \nVdash \theta_2$, we can infer that it would have to be that $B(1) \cap \{-n \mid n \in \omega\}$ has exactly one member (since $\{-n \mid n \in \omega\}$ is the set of objects that do not exist at 0). But if that is right, then these models are distinguishable by the following $\mathcal{L}^{1M}(\approx, \Pi)$-formula:

$$\Sigma x \, \Sigma y \, (x \napprox y \wedge \neg \mathsf{E}(x) \wedge \neg \mathsf{E}(y) \wedge \Diamond(\mathsf{E}(x) \wedge \neg \mathsf{E}(y))).$$

The proof of **Lemma D.4** was inspired by an attempt to fix Hodes's proof.

Proof: It is easy to show that if φ is an $\mathcal{L}^{1M}(\approx, \downarrow, \mathcal{F}, \Pi)$-formula, then $\Vdash_{\mathbf{U}} \mathcal{F}\varphi \leftrightarrow \varphi$ (just use the observation from the proof of **Lemma D.1**). So every $\mathcal{L}^{1M}(\approx, \downarrow, \mathcal{F}, \Pi)$-formula is **U**-equivalent to an $\mathcal{L}^{1M}(\approx, \downarrow, \Pi)$-formula. But if ψ is an $\mathcal{L}^{1M}(\approx, \downarrow, \Pi)$-formula, then $\Vdash_{\mathbf{U}} \downarrow\psi \leftrightarrow \psi$. Putting these together, every $\mathcal{L}^{1M}(\approx, \downarrow, \mathcal{F}, \Pi)$-formula is **U**-equivalent to an $\mathcal{L}^{1M}(\approx, \Pi)$-formula. So it suffices to find two $\mathcal{L}^{1M}(\approx, \Pi)$-bisimilar models in **UD** that disagree on some $\mathcal{L}^{1M}(@)$-formula.

Let $\mathcal{M}_1 = \langle W_1, R_1, F_1, D_1, \delta_1, I_1 \rangle$, where:

$$W_1 = \{w\} \cup \{v_S^T \mid S \subseteq \mathbb{N}, T \subseteq \mathbb{N}^-, \text{ and } 1 < |S|, |T| < \aleph_0\},$$

R_1 and F_1 are universal, $D_1 = \mathbb{Z}$, $\delta_1(w) = \mathbb{N}$, $\delta_1(v_S) = (\mathbb{N} - S) \cup T$, and $I_1 = \varnothing$. Let $\mathcal{M}_2$ be like $\mathcal{M}_1$ except that $W_2 = W_1 \cup \{v\}$, and $\delta_2(v) = (\mathbb{N} - \{1\}) \cup \{-1\}$. Observe:

$$\mathcal{M}_1, w, w \Vdash \Box\,(\exists x\, @\,\neg\,\mathsf{E}(x) \rightarrow \exists_{\geqslant 2} x\, @\,\neg\,\mathsf{E}(x))$$
$$\mathcal{M}_2, w, w \nVdash \Box\,(\exists x\, @\,\neg\,\mathsf{E}(x) \rightarrow \exists_{\geqslant 2} x\, @\,\neg\,\mathsf{E}(x)).$$

However, we will show that $\mathcal{M}_1, w, w \leftrightarroweq_{\approx, \Pi} \mathcal{M}_2, w, w$. Clearly $w, w \simeq w, w$. Suppose $w, u_1, \bar{a} \simeq w, u_2, \bar{b}$. Since $\delta_1(u_1)$ and $\overline{\delta_1(u_1)}$ are infinite, if Abelard picks an $a' \in D_1$, then Eloïse can find a $b' \in D_2$ so that $w, u_1, \bar{a}, a' \simeq w, u_2, \bar{b}, b'$ by ensuring that $a' \in \delta_1(u_1)$ iff $b' \in \delta_2(u_2)$. Likewise if Abelard picks a $b' \in D_2$. Now, suppose Abelard picks an $u_1' \in W_1$. Define $S = \{b_i \in \mathbb{N} \mid a_i \notin \delta_1(u_1)\}$ and $T = \{b_i \in \mathbb{N}^- \mid a_i \in \delta_1(u_1)\}$. (If $|S| \leqslant 1$, add a couple of elements from $\mathbb{N} - \{\bar{b}\}$ to S. If $|T| \leqslant 1$, add a couple of elements from $\mathbb{N}^- - \{\bar{b}\}$ to T.) Then $a_i \in \delta_1(u_1)$ iff $b_i \in \delta_2(v_S^T)$. So $w, v_S^T, \bar{a} \simeq w, v_{S'}^{T'}, \bar{b}$. Likewise if Abelard chooses a $u_2' \in W_2$, even if $u_2' = v$. Thus, using **Lemma D.3**, $\mathcal{L}^{1M}(@) \not\leqslant_{\mathbf{UD}} \mathcal{L}^{1M}(\approx, \Pi)$. ∎

Lemma D.5 (*Adding Two Operators*). $\mathcal{L}^{1M}(\downarrow, \mathcal{F}) \not\leqslant_{\mathbf{D}} \mathcal{L}^{1M}(\approx, @, \mathcal{F}, \Pi) \not\geqslant_{\mathbf{D}} \mathcal{L}^{1M}(@, \downarrow)$.

Proof: First, $\mathcal{L}^{1M}(\downarrow, \mathcal{F}) \not\leqslant_{\mathbf{D}} \mathcal{L}^{1M}(@, \mathcal{F})$. Let $\mathcal{M}_1 = \langle W_1, R_1, F_1, D_1, \delta_1, I_1 \rangle$, where $W_1 = \{w, v\}$, $R_1 = W_1 \times W_1$, $F_1 = \varnothing$, $D_1 = \delta_1(w) = \delta_1(v) = \{a\}$, and $I_1 = \varnothing$. Let $\mathcal{M}_2$ be like $\mathcal{M}_1$ except $F_2 = \{\langle v, v \rangle\}$. Then $\mathcal{M}_1, w, w \leftrightarroweq_{\approx, @, \mathcal{F}, \Pi} \mathcal{M}_2, w, w$ (since $F_1[w] = F_2[w] = \varnothing$), but $\mathcal{M}_1, w, w \Vdash \Box\downarrow\mathcal{F}\bot$, while $\mathcal{M}_2, w, w \nVdash \Box\downarrow\mathcal{F}\bot$. So $\mathcal{L}^{1M}(\downarrow, \mathcal{F}) \not\leqslant_{\mathbf{D}} \mathcal{L}^{1M}(\approx, @, \mathcal{F}, \Pi)$.

Next, $\mathcal{L}^{1M}(@, \downarrow) \not\leqslant_{\mathbf{D}} \mathcal{L}^{1M}(\approx, @, \mathcal{F}, \Pi)$. Consider the models $\mathcal{N}_1$ and $\mathcal{N}_2$ from Figure 4. Modify them so that $F_1 = F_2 = \varnothing$, and call the resulting models $\mathcal{N}_1'$ and $\mathcal{N}_2'$. Then $\mathcal{N}_1', z, z \leftrightarroweq_{\approx, @, \mathcal{F}, \Pi} \mathcal{N}_2', z, z$, but they disagree on $\Box\downarrow\Diamond\forall x\,(@\mathsf{Rich}(x) \rightarrow \mathsf{Poor}(x))$. Hence, $\mathcal{L}^{1M}(@, \downarrow) \not\leqslant_{\mathbf{D}} \mathcal{L}^{1M}(\approx, @, \mathcal{F}, \Pi)$. ∎

It is tedious, but straightforward, to show the following using the lemmas above:

Theorem D.6 (*Completeness of Figure 8*). Figure 8 is a complete diagram of the expressive power of the languages presented in that diagram.

Now we turn to extensions with E, $\approx$, and Π. It will help to define back-and-forth games for $\mathcal{L}^{\mathsf{TS}}$ and some of its fragments. First, an $\mathcal{L}^{\mathsf{TS}}$-formula is **almost E-free** if it can be built from atomics other than those of the form $\mathsf{E}(x;s)$ using negation, conjunction, object quantification, state quantification, and E-bounded object quantification. (Thus, E only occurs as the bounds of object quantifiers.) Let $\mathcal{L}^{\mathsf{TS}}_{-\approx}$ be the $\approx$-free fragment, $\mathcal{L}^{\mathsf{TS}}_{-\approx,(\mathsf{E})}$ be the $\approx$-free and almost E-free fragment, and $\mathcal{L}^{\mathsf{TS}}_{\restriction\mathsf{E}}$ be the E-bounded fragment of $\mathcal{L}^{\mathsf{TS}}$.

Definition D.7 *(Back-and-Forth System).* Let $\mathcal{M}$ and $\mathcal{N}$ be models. A **back-and-forth system between $\mathcal{M}$ and $\mathcal{N}$** is a nonempty variably polyadic relation Z such that whenever $Z(\overline{w}, \overline{a}; \overline{v}, \overline{b})$, $|\overline{w}| = |\overline{v}| < \aleph_0$ and $|\overline{a}| = |\overline{b}| < \aleph_0$, and if $Z(\overline{w}, \overline{a}; \overline{v}, \overline{b})$, then:

(TS Atomic) $\forall k \leqslant |\overline{w}| \ \forall m \in \mathbb{N} \ \forall P^m \in \mathsf{PRED}^m \ \forall i_1, \ldots, i_m \leqslant |\overline{a}|$:

$$\langle a_{i_1}, \ldots, a_{i_m} \rangle \in I^{\mathcal{M}}(P^m, w_k) \quad \Leftrightarrow \quad \langle b_{i_1}, \ldots, b_{i_m} \rangle \in I^{\mathcal{N}}(P^m, v_k)$$

(TS Eq) $\forall n, m \leqslant |\overline{a}| : a_n = a_m$ iff $b_n = b_m$

(TS StEq) $\forall k, l \leqslant |\overline{w}| : w_k = w_l$ iff $v_k = v_l$

(TS Ex) $\forall k \leqslant |\overline{w}| \ \forall n \leqslant |\overline{a}| : a_n \in \delta^{\mathcal{M}}(w_k)$ iff $b_n \in \delta^{\mathcal{N}}(v_k)$

(TS Acc) $\forall k, l \leqslant |\overline{w}| : R(w_k, w_l)$ iff $R(v_k, v_l)$ and $F(w_k, w_l)$ iff $F(v_k, v_l)$

(TS Zig) $\forall w' \in W^{\mathcal{M}} \exists v' \in W^{\mathcal{N}} : Z(\overline{w}, w', \overline{a}; \overline{v}, v', \overline{b})$

(TS Zag) $\forall v' \in W^{\mathcal{N}} \exists w' \in W^{\mathcal{M}} : Z(\overline{w}, w', \overline{a}; \overline{v}, v', \overline{b})$

(TS Forth) $\forall a' \in D^{\mathcal{M}} \exists b' \in D^{\mathcal{N}} : Z(\overline{w}, \overline{a}, a'; \overline{v}, \overline{b}, b')$

(TS Back) $\forall b' \in D^{\mathcal{N}} \exists a' \in D^{\mathcal{M}} : Z(\overline{w}, \overline{a}, a'; \overline{v}, \overline{b}, b')$.

We may write $\mathcal{M}, \overline{w}, \overline{a} \leftrightarrows_{\mathsf{TS}} \mathcal{N}, \overline{v}, \overline{b}$ to indicate that $\mathcal{M}, \overline{w}, \overline{a}$ and $\mathcal{N}, \overline{v}, \overline{b}$ are **back-and-forth equivalent**. If we drop (TS Eq) and (TS StEq), we get a notion of a back-and-forth system for $\mathcal{L}^{\mathsf{TS}}_{-\approx}$. We get a notion of a back-and-forth system for $\mathcal{L}^{\mathsf{TS}}_{-\approx,(\mathsf{E})}$ if we drop (TS Eq), (TS StEq), and (TS Ex) and we add:

(TS E-Forth) $\forall k \leqslant |\overline{w}| \ \forall a' \in \delta^{\mathcal{M}}(w_k) \exists b' \in \delta^{\mathcal{N}}(v_k) : Z(\overline{w}, \overline{a}, a'; \overline{v}, \overline{b}, b')$

(TS E-Back) $\forall k \leqslant |\overline{v}| \ \forall b' \in \delta^{\mathcal{N}}(v_k) \exists a' \in \delta^{\mathcal{M}}(w_k) : Z(\overline{w}, \overline{a}, a'; \overline{v}, \overline{b}, b')$.

If we replace (TS Forth) and (TS Back) with (TS E-Forth) and (TS E-Back), we get a notion of a back-and-forth system for $\mathcal{L}^{\mathsf{TS}}_{\restriction\mathsf{E}}$.

Definition D.8 *($\mathcal{L}^{\mathsf{TS}}$-Equivalence).* $\mathcal{M}, \overline{w}, \overline{a}$ and $\mathcal{N}, \overline{v}, \overline{b}$ are **$\mathcal{L}^{\mathsf{TS}}$-equivalent** if for all $\mathcal{L}^{\mathsf{TS}}$-formulas $\alpha(\overline{x}; \overline{s})$ (where $|\overline{x}| \leqslant |\overline{a}|$ and $|\overline{s}| \leqslant |\overline{w}|$), $\mathcal{M} \vDash \alpha[\overline{a}; \overline{w}]$ iff $\mathcal{N} \vDash \alpha[\overline{b}; \overline{v}]$. We may write "$\mathcal{M}, \overline{w}, \overline{a} \equiv_{\mathsf{TS}} \mathcal{N}, \overline{v}, \overline{b}$" to indicate that $\mathcal{M}, \overline{w}, \overline{a}$ and $\mathcal{N}, \overline{v}, \overline{b}$ are $\mathcal{L}^{\mathsf{TS}}$-equivalent. Likewise for the various fragments of $\mathcal{L}^{\mathsf{TS}}$.

It is easy to check that $\mathcal{M}, \overline{w}, \overline{a} \leftrightarrows_{\mathsf{TS}} \mathcal{N}, \overline{v}, \overline{b}$ implies $\mathcal{M}, \overline{w}, \overline{a} \equiv_{\mathsf{TS}} \mathcal{N}, \overline{v}, \overline{b}$, and likewise for the various fragments of $\mathcal{L}^{\mathsf{TS}}$. Now, say $\mathcal{L}_1 \leqslant^* \mathcal{L}_2$ if every $\mathcal{L}_1$-formula is equivalent to some $\mathcal{L}_2$-formula. This is more stringent than $\leqslant$, since some $\mathcal{L}_1$-formula might only be expressible as a set of $\mathcal{L}_2$-formulas. Observe by **Definition 2.6** that $\mathcal{L}^{\mathsf{1M}}(@, \downarrow, \mathcal{F}, \Pi) \leqslant^* \mathcal{L}^{\mathsf{TS}}_{-\approx,(\mathsf{E})}$, that $\mathcal{L}^{\mathsf{1M}}(\mathsf{E}, @, \downarrow, \mathcal{F}, \Pi) \leqslant^* \mathcal{L}^{\mathsf{TS}}_{-\approx}$, and that $\mathcal{L}^{\mathsf{1M}}(\approx, @, \downarrow, \mathcal{F}) \leqslant^* \mathcal{L}^{\mathsf{TS}}_{\upharpoonright \mathsf{E}}$.

> **Lemma D.9** (*Adding E and* $\approx$). If $\mathcal{L}^{\mathsf{1M}} \leqslant^* \mathcal{L} \leqslant^* \mathcal{L}^{\mathsf{TS}}_{-\approx,(\mathsf{E})}$, then $\mathcal{L} < \mathcal{L}(\mathsf{E}) < \mathcal{L}(\approx)$. Likewise if relativize to **D**, **U**, or **UD**.

> **Proof**: Recall the models $\mathcal{E}$ and $\mathcal{E}'$ from Figure 1. It is easy to check that via our original bisimulation, $\mathcal{E}, w, w \leftrightarrows_{\mathsf{TS}-\approx,(\mathsf{E})} \mathcal{E}', w, w$ (remember, you do not need to satisfy (TS Eq) or (TS Ex) in this back-and-forth game!). But these models are distinguishable by the $\mathcal{L}^{\mathsf{1M}}(\mathsf{E})$-formula $\Diamond \exists x \Diamond \neg \mathsf{E}(x)$. So $\mathcal{L}^{\mathsf{1M}}(\mathsf{E}) \not\leqslant \mathcal{L}^{\mathsf{TS}}_{-\approx,(\mathsf{E})}$. Suppose now for *reductio* that $\mathcal{L}(\mathsf{E}) \leqslant \mathcal{L}$. Since $\mathcal{L}^{\mathsf{1M}}(\mathsf{E}) \leqslant^* \mathcal{L}(\mathsf{E})$ (easily verified by induction), $\mathcal{L}^{\mathsf{1M}}(\mathsf{E}) \leqslant \mathcal{L} \leqslant^* \mathcal{L}^{\mathsf{TS}}_{-\approx,(\mathsf{E})}$, ↯. So $\mathcal{L}(\mathsf{E}) \not\leqslant \mathcal{L}$, and thus $\mathcal{L} < \mathcal{L}(\mathsf{E})$.
>   As for $\mathcal{L}(\mathsf{E}) < \mathcal{L}(\approx)$, revise $\mathcal{E}$ and $\mathcal{E}'$ by deleting the world w from the models. Call the resulting models $\mathcal{E}_-$ and $\mathcal{E}'_-$. Then $\mathcal{E}_-, v \leftrightarrows_{\mathsf{TS}-\approx} \mathcal{E}'_-, v$, but they disagree on $\exists x \exists y \, (x \not\approx y)$. So $\mathcal{L}^{\mathsf{1M}}(\approx) \not\leqslant \mathcal{L}^{\mathsf{TS}}_{-\approx}$. But $\mathcal{L}(\mathsf{E}) \leqslant \mathcal{L}(\approx)$, so reasoning as before (noting that $\mathcal{L}(\mathsf{E}) \leqslant^* \mathcal{L}^{\mathsf{TS}}_{-\approx}$), we have that $\mathcal{L}(\mathsf{E}) < \mathcal{L}(\approx)$.   ∎

Now, where $\mathcal{L}_1$ and $\mathcal{L}_2$ were languages in Figure 8 such that $\mathcal{L}_1 < \mathcal{L}_2$, we can show that the inclusions involving their extensions with E or $\approx$ can be diagrammed as in Figure 9. First, the arrows that are present are immediate by **Lemma D.9** and by the fact that if $\mathcal{L}_1 < \mathcal{L}_2$ in Figure 8, then we already have $\mathcal{L}_1 <^* \mathcal{L}_2$. Next, $\mathcal{L}_1(\mathsf{E}) \not\leqslant \mathcal{L}_2$, since if it were, we would have $\mathcal{L}^{\mathsf{1M}}(\mathsf{E}) \leqslant^* \mathcal{L}_1(\mathsf{E}) \leqslant \mathcal{L}_2 \leqslant^* \mathcal{L}^{\mathsf{TS}}_{-\approx,(\mathsf{E})} \not\geqslant \mathcal{L}^{\mathsf{1M}}(\mathsf{E})$, contrary to **Lemma D.9**. Likewise, $\mathcal{L}_1(\approx) \not\leqslant \mathcal{L}_2(\mathsf{E})$. Finally, observe that in the results used above to show that $\mathcal{L}_2 \not\leqslant \mathcal{L}_1$, we already showed that $\mathcal{L}_2 \not\leqslant \mathcal{L}_1(\approx, \Pi)$. Thus, $\mathcal{L}_2 \not\leqslant \mathcal{L}_1(\approx)$.
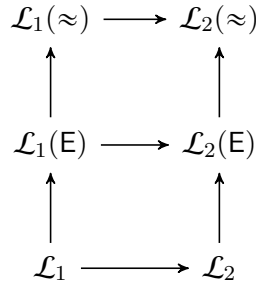
$$\begin{array}{ccc}
\mathcal{L}_1(\approx) & \longrightarrow & \mathcal{L}_2(\approx) \\
\uparrow & & \uparrow \\
\mathcal{L}_1(\mathsf{E}) & \longrightarrow & \mathcal{L}_2(\mathsf{E}) \\
\uparrow & & \uparrow \\
\mathcal{L}_1 & \longrightarrow & \mathcal{L}_2
\end{array}$$

Figure 9: Relative expressive power after adding E or $\approx$.

Lemma D.10 (*Adding* Π). If $\mathcal{L}^{1M} \leqslant^* \mathcal{L} \leqslant^* \mathcal{L}^{TS}_{\restriction E}$, then $\mathcal{L} < \mathcal{L}(\Pi)$. Likewise if we relativize to **U**. Also, if $\mathcal{L}^{1M} \leqslant^* \mathcal{L} \leqslant^* \mathcal{L}^{1M}(\approx, @, \downarrow, \mathcal{F})$, then $\mathcal{L} <_\mathbf{D} \mathcal{L}(\Pi)$.

Proof: Let $\mathcal{M}_1 = \langle W_1, R_1, F_1, D_1, \delta_1, I_1 \rangle$, where $W_1 = \{w\}$, $R_1 = F_1 = \{\langle w, w \rangle\}$, $D_1 = \{a\}$, $\delta_1(w) = \{a\}$, $I_1(P, w) = \{a\}$. Let $\mathcal{M}_2$ be just like $\mathcal{M}_1$ except $D_2 = \{a, b\}$. Then $\mathcal{M}_1, w \leftrightarroweq_{TS\restriction E} \mathcal{M}_2, w$, but they disagree on the $\mathcal{L}^{1M}(\Pi)$-formula $\Sigma x \neg P(x)$. So $\mathcal{L}^{1M}(\Pi) \not\leqslant \mathcal{L}^{TS}_{\restriction E}$. Reasoning as before, $\mathcal{L} < \mathcal{L}(\Pi)$.

Suppose now we restrict to **D**. Let $\mathcal{M}_1 = \langle W_1, R_1, F_1, D_1, \delta_1, I_1 \rangle$, where $W_1 = \{w\} \cup \{v_n \mid n \in \mathbb{N}\}$, $R_1 = F_1 = \{\langle w, v_n \rangle \mid n \in \mathbb{N}\}$, $\delta_1(w) = \mathbb{N}$, $\delta_1(v_n) = \mathbb{N} - \{n\}$, $I(P, w) = \varnothing$, and for each $n \in \mathbb{N}$, $I(P, v_n) = \{n\}$. Let $\mathcal{M}_2$ be just like $\mathcal{M}_1$ except $W_2 = W_1 \cup \{u\}$, where $u \notin W_1$, $R_2 = R_1 \cup \{\langle w, u \rangle\}$, $F_2 = F_1$, $\delta_2(u) = \mathbb{N}$, and $I(P, u) = \varnothing$. One can show that $\mathcal{M}_1, w, w \leftrightarroweq_{\approx, @, \downarrow, \mathcal{F}} \mathcal{M}_2, w, w$. But, they disagree on the $\mathcal{L}^{1M}(\Pi)$-formula $\Box \Sigma x\, P(x)$. So $\mathcal{L}^{1M}(\Pi) \not\leqslant_\mathbf{D} \mathcal{L}^{1M}(\approx, @, \downarrow, \mathcal{F})$. So $\mathcal{L} <_\mathbf{D} \mathcal{L}(\Pi)$.   ∎

So once again, using **Lemma D.10** and the results above, we can verify that if $\mathcal{L}_1$ and $\mathcal{L}_2$ are in Figure 8 and $\mathcal{L}_1 < \mathcal{L}_2$, then their extensions involving Π can be represented in Figure 10. This holds even if we add E or $\approx$. Thus, Figure 5 from §4 is correct. Moreover, it is still correct even relative to **D**.
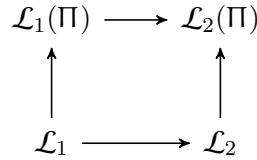
$$\mathcal{L}_1(\Pi) \longrightarrow \mathcal{L}_2(\Pi)$$
$$\uparrow \qquad\qquad \uparrow$$
$$\mathcal{L}_1 \longrightarrow \mathcal{L}_2$$

Figure 10: Relative expressive power after adding Π.

We now turn to asking to what extent these results hold relative to **U** and **UD**. We only give a partial answer here. First, set aside E, $\approx$, and Π, and focus just on **U**. Then the diagram of expressive power looks something like Figure 11 (whether we should include the dashed arrows has yet to be determined).

First, if φ is @-free, then $\Vdash_\mathbf{U} (\mathcal{F}\varphi \leftrightarrow \varphi)$ and $\Vdash_\mathbf{U} (\downarrow\varphi \leftrightarrow \varphi)$. So $\mathcal{L}^{1M} \equiv_\mathbf{U} \mathcal{L}^{1M}(\downarrow) \equiv_\mathbf{U} \mathcal{L}^{1M}(\mathcal{F}) \equiv_\mathbf{U} \mathcal{L}^{1M}(\downarrow, \mathcal{F})$. But still $\mathcal{L}^{1M} <_\mathbf{U} \mathcal{L}^{1M}(@)$ by **Lemma D.4**. And the remarks on page 16 (together with **Lemma D.3**) show that $\mathcal{L}^{1M}(@) <_\mathbf{U} \mathcal{L}^{1M}(@, \downarrow)$ and that $\mathcal{L}^{1M}(@) <_\mathbf{U} \mathcal{L}^{1M}(@, \mathcal{F})$. As for the lack of inclusion from $\mathcal{L}^{1M}(@, \mathcal{F})$ to $\mathcal{L}^{1M}(@, \downarrow)$:

Lemma D.11 (*$\mathcal{F}, @$ Not Included in $@, \downarrow$*). $\mathcal{L}^{1M}(@, \mathcal{F}) \not\leqslant_\mathbf{UD} \mathcal{L}^{1M}(\approx, @, \downarrow)$.

Proof: Let $\mathcal{M}_1 = \langle W_1, R_1, F_1, D_1, \delta_1, I_1 \rangle$ where:

$$W_1 = \{v_S^T \mid S \subseteq \mathbb{N}, T \subseteq \mathbb{N}^-, |T| < \aleph_0, \text{ and either } S = \varnothing \text{ or } 1 < |S| < \aleph_0\},$$
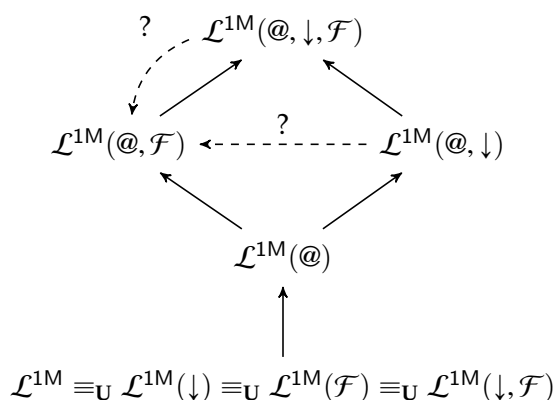
Figure 11: Relative **U**-expressive power for languages between $\mathcal{L}^{1M}$ and $\mathcal{L}^{1M}(@,\downarrow,\mathcal{F})$.

$R_1 = F_1 = W_1 \times W_1$, $D_1 = \mathbb{Z}$, $\delta_1(w) = \mathbb{N}$, $\delta_1(v_S^T) = (\mathbb{N} - S) \cup T$, and $I_1 = \varnothing$. Let $\mathcal{M}_2$ be just like $\mathcal{M}_1$, except we allow $|S| = 1$. Let $w = v_\varnothing^\varnothing$. Observe that:

$$\mathcal{M}_1, w, w \Vdash \mathcal{F}\, (\exists x\, @\neg\, \mathsf{E}(x) \rightarrow \exists_{\geqslant 2} x\, @\neg\, \mathsf{E}(x))$$
$$\mathcal{M}_2, w, w \nVdash \mathcal{F}\, (\exists x\, @\neg\, \mathsf{E}(x) \rightarrow \exists_{\geqslant 2} x\, @\neg\, \mathsf{E}(x)).$$

However, we will show $\mathcal{M}_1, w, w \leftrightarrows_{\approx,@,\downarrow} \mathcal{M}_2, w, w$. Clearly $w, w \simeq w, w$. Suppose throughout that $u_1, u_2, \overline{a} \simeq u_1', u_2', \overline{b}$ and that the following hold:

(I) $a_i \in \delta_1(u_1)$ iff $b_i \in \delta_1(u_1)$

(II) $u_1 = u_2$ iff $u_1' = u_2'$

(III) $\left|(\overline{\delta_1(u_1)} \cap \delta_1(u_2)) - \{\overline{a}\}\right| = \left|(\overline{\delta_2(u_1')} \cap \delta_2(u_2')) - \{\overline{b}\}\right|$.

Observe that no matter what u_1 and u_2 are, $(\delta_1(u_1) \cap \delta_1(u_2)) - \{\overline{a}\}$ is infinite, and $\overline{(\delta_1(u_1)} \cap \delta_1(u_2)) - \{\overline{a}\}$ is finite. Likewise for u_1' and u_2'.

First, suppose Abelard picks a new $a \in \delta_1(u_2)$. If $a \in \delta_1(u_1)$, then since $(\delta_2(u_1') \cap \delta_2(u_2')) - \{\overline{b}\}$ is infinite, Eloïse will always be able to match a with a $b \in (\delta_2(u_1') \cap \delta_2(u_2')) - \{\overline{b}\}$. If instead $a \notin \delta_1(u_1)$, then by (III), we can find a $b \in (\overline{\delta_2(u_1')} \cap \delta_2(u_2')) - \{\overline{b}\}$ to match a with. A symmetric argument applies if Abelard instead picks a $b \in \delta_2(u_2')$.

Next, suppose Abelard decides to relocate the game. If he invokes (Act) or (Diag), then clearly (I)–(III) hold. So suppose he decides to pick a $u_3 \in W_3$ to relocate to. Eloïse's choice is obvious if $u_3 = u_1$, so suppose $u_3 \neq u_1$. We will construct a T_3' and a S_3' of the appropriate sort and show they meet (I)–(III). First, pick two elements $c, d \in (\mathbb{N} \cap \delta_2(u_1')) - \{\overline{b}\}$ (note that $(\mathbb{N} \cap \delta_2(u_1')) - \{\overline{b}\}$ is infinite since each world has cofinitely many positive integers) and define:

$$S_3' := \{b_i \in \mathbb{N} \mid a_i \notin \delta_1(u_3)\} \cup (\mathbb{N} - (\delta_2(u_1') \cup \{\overline{b}\})) \cup \{c, d\}.$$

41

Note that where $u_1' = v_{S_1'}^{T_1'}$, $\mathbb{N} - (\delta_2(u_1') \cup \{\overline{b}\}) = S_1' - \{\overline{b}\}$ is finite, so S_3' is finite and $S_1' - \{\overline{b}\} \subseteq S_3'$. Second, define $T_{3,0}' := \{b_i \in \mathbb{N}^- \mid a_i \in \delta_1(u_3)\}$. Observe that:

$$[\overline{\delta_2(u_1')} \cap \delta_2(v_{S_3'}^{T_{3,0}'})] - \{\overline{b}\} = [(S_1' \cup (\mathbb{N}^- - T_1')) \cap ((\mathbb{N} - S_3') \cup T_{3,0}')] - \{\overline{b}\}$$
$$= [(S_1' \cap (\mathbb{N} - S_3')) \cup ((\mathbb{N}^- - T_1') \cap T_{3,0}')] - \{\overline{b}\}$$
$$= [(S_1' \cap \{\overline{b}\}) \cup T_{3,0}'] - \{\overline{b}\} = \varnothing.$$

Now, where $k = |(\overline{\delta_1(u_1)} \cap \delta_1(u_2)) - \{\overline{a}\}|$, pick k-many elements $e_1, \ldots, e_k \in \mathbb{N}^- - (\delta_2(u_1') \cup \{\overline{b}\})$ (notice that $\mathbb{N}^- - (\delta_2(u_1') \cup \{\overline{b}\})$ is infinite, since each world only has finitely many negative integers). Define $T_3' = T_{3,0}' \cup \{e_1, \ldots, e_k\}$. We will show that if Eloïse chooses $u_3' = v_{S_3'}^{T_3'}$, then all the necessary constraints are met.

We first need to show $u_1, u_3, \overline{a} \simeq u_1', v_{S_3'}^{T_3'}, \overline{b}$—in particular, $a_i \in \delta_1(u_3)$ iff $b_i \in \delta_2(u_3')$. Suppose $a_i \in \delta_1(u_3)$. Either $b_i \in \mathbb{N}$ or $b_i \in \mathbb{N}^-$. If $b_i \in \mathbb{N}$, then $b_i \notin S_3'$, so $b_i \in \delta_2(u_3')$. If $b_i \in \mathbb{N}^-$, then $b_i \in T_{3,0}' \subseteq T_3'$, so $b_i \in \delta_2(u_3')$. Suppose instead $a_i \notin \delta_1(u_3)$. Again, either $b_i \in \mathbb{N}$ or $b_i \in \mathbb{N}^-$. If $b_i \in \mathbb{N}$, then $b_i \in S_3'$, so $b_i \notin \delta_2(u_3')$. If $b_i \in \mathbb{N}^-$, then $b_i \notin T_3'$, so $b_i \notin \delta_2(u_3')$. No matter what, $a_i \in \delta_1(u_3)$ iff $b_i \in \delta_2(u_3')$.

Next, we need to show (I)–(III). (I) holds by default. Now, we assumed above $u_3 \neq u_1$, so we need $u_3' \neq u_1'$. But recall that we picked c, d so that $c, d \in \delta_2(u_1')$. But $c, d \in S_3'$, so $c, d \notin \delta_2(u_3')$. Thus, $u_3' \neq u_1'$. So (II) holds. Finally, using the calculations above, since $e_1, \ldots, e_k \in \mathbb{N}^- - (\delta_2(u_1') \cup \{\overline{b}\}) = \mathbb{N}^- - (T_1' \cup \{\overline{b}\})$, we find that:

$$[\overline{\delta_2(u_1')} \cap \delta_2(v_{S_3'}^{T_{3,0}'})] - \{\overline{b}\} = [(S_1' \cap \{\overline{b}\}) \cup (T_{3,0}' \cup \{e_1, \ldots, e_k\})] - \{\overline{b}\}$$
$$= \{e_1, \ldots, e_k\},$$

where $k = |(\overline{\delta_1(u_1)} \cap \delta_1(u_2)) - \{\overline{a}\}|$. So (III) holds. Thus, if Abelard relocates to u_3, Eloïse can choose to relocate to u_3'. And since $|S_3'| > 1$, a symmetric argument applies if Abelard decides to relocate the game in $\mathcal{M}_2$. The proof is completed with one application of **Lemma D.3**. ∎

Now, because the proof of **Lemma D.9** only uses models in **U** (in fact, in **UD**), we can still safely say that adding E or $\approx$ can be diagrammed as in Figure 9. Adding Π makes things more complicated. Recall that relative to the class of all models, we could simply say that if $\mathcal{L}^{1M} <^* \mathcal{L}_1 <^* \mathcal{L}_2 \leqslant^* \mathcal{L}^{1M}(\approx, @, \downarrow, \mathcal{F})$, then $\mathcal{L}_i < \mathcal{L}_i(\Pi)$, $\mathcal{L}_1(\Pi) < \mathcal{L}_2(\Pi)$, and $\mathcal{L}_1(\Pi)$ and $\mathcal{L}_2$ were incomparable. However, showing that $\mathcal{L}_2 \nleqslant \mathcal{L}_1(\Pi)$ crucially relied on the fact that all of our inexpressibility proofs for showing $\mathcal{L}_2 \nleqslant \mathcal{L}_1$ already showed that $\mathcal{L}_2 \nleqslant \mathcal{L}_1(\Pi)$. But because **Lemma D.11** left out Π (which is crucial, as we will see below), we cannot conclude that $\mathcal{L}_2 \nleqslant_{\mathbf{U}} \mathcal{L}_1(\Pi)$.

We can still verify by hand that in some cases, $\mathcal{L}_2 \nleqslant \mathcal{L}_1(\Pi)$. For one thing, $\mathcal{L}^{1M}(@) \nleqslant_{\mathbf{UD}} \mathcal{L}^{1M}(\approx, \Pi)$ by **Lemma D.4**. We also have that $\mathcal{L}^{1M}(@, \downarrow) \nleqslant_{\mathbf{UD}} \mathcal{L}^{1M}(\approx, @, \Pi)$ by **Proposition 4.4**. Likewise, $\mathcal{L}^{1M}(@, \mathcal{F}) \nleqslant_{\mathbf{UD}} \mathcal{L}^{1M}(\approx, @, \Pi)$. But importantly, some of these languages without E and Π that were distinct collapse when you add E and Π:

Lemma D.12 *(Collapse).* $\mathcal{L}^{1M}(\mathsf{E}, @, \downarrow, \Pi) \equiv_{\mathbf{U}} \mathcal{L}^{1M}(\mathsf{E}, @, \mathcal{F}, \Pi) \equiv_{\mathbf{U}} \mathcal{L}^{1M}(\mathsf{E}, @, \downarrow, \mathcal{F}, \Pi)$. Likewise if we add $\approx$ to these languages.

Proof: Throughout, let $\mathcal{L}^* = \mathcal{L}^{1M}(\mathsf{E}, @, \downarrow, \mathcal{F}, \Pi)$. Note that the following are all **U**-valid (where α is an atomic formula):

$$\downarrow \alpha \leftrightarrow \alpha$$
$$\downarrow \neg \varphi \leftrightarrow \neg \downarrow \varphi$$
$$\downarrow (\varphi \wedge \psi) \leftrightarrow (\downarrow \varphi \wedge \downarrow \psi)$$
$$\downarrow @ \varphi \leftrightarrow \downarrow \varphi$$
$$\downarrow \downarrow \varphi \leftrightarrow \downarrow \varphi$$
$$\downarrow \mathcal{F} \varphi \leftrightarrow \mathcal{F} \varphi$$
$$\downarrow \Pi x \varphi \leftrightarrow \Pi x \downarrow \varphi.$$

Likewise, all of these are **U**-valid:

$$@ \neg \varphi \leftrightarrow \neg @ \varphi$$
$$@(\varphi \wedge \psi) \leftrightarrow (@\varphi \wedge @\psi)$$
$$@\Box \varphi \leftrightarrow \Box \varphi$$
$$@@\varphi \leftrightarrow @\varphi$$
$$@\downarrow \varphi \leftrightarrow @\varphi$$
$$@\Pi x \varphi \leftrightarrow \Pi x @\varphi.$$

Using these rules, we can push each $@$ and each $\downarrow$ inwards as much as possible until $@$ only occurs right before a $\mathcal{F}$ or an atomic, and $\downarrow$ only occurs right before a $\Box$. Moreover, we can delete any $\mathcal{F}$ and $\downarrow$ if it does not scope over an $@$, and repeat. After this entire process, the resulting formula will be **U**-equivalent to our original. So assume without loss of generality that our formula has already gone through this pre-processing.

Now, say that an $\mathcal{L}^*$-formula is in ***normal form*** if it is either a non-modal formula, or if it is of the form:

$$Q_1 y_1 \cdots Q_n y_n \mathsf{BC}(\overline{\psi}, \overline{\star \theta}),$$

where $Q_i \in \{\Sigma, \Pi\}$ (the quantifier block may be empty), BC is some boolean combination of its components, $\overline{\psi}$ are all non-modal, each $\star_i \in \{\Box, @, \downarrow\Box, \mathcal{F}\}$, and $\overline{\theta}$ are all in normal form. By induction, one can convert every $\mathcal{L}^*$-formula into one of normal form (essentially by pre-processing as above, and then replacing bound variables and pulling out quantifiers). Thus, we may assume without loss of generality that our formula is already in normal form.

Finally, suppose an $\mathcal{L}^*$-formula has been pre-processed and is in the form:

$$Q_1 y_n \cdots Q_n y_n \mathsf{BC}(\overline{\varphi}, \overline{@\psi}, \overline{\Box\theta}, \overline{\downarrow\Box\chi}, \overline{\mathcal{F}\xi}),$$

where $\overline{\varphi}$ are all non-modal, and $\overline{\psi}, \overline{\theta}, \overline{\chi}$, and $\overline{\xi}$ are all in normal form (notice that since we pre-processed, each ψ is either an atomic or of the form $\mathcal{F}\psi'$). Then the following are **U**-valid:

$$\mathcal{F}Q_1 y_n \cdots Q_n y_n \mathrm{BC}(\overline{\varphi}, \overline{@\psi}, \overline{\Box\theta}, \overline{\downarrow\Box\chi}, \overline{\mathcal{F}\xi}) \leftrightarrow \downarrow\Box Q_1 y_n \cdots Q_n y_n \mathrm{BC}(\overline{@\varphi}, \overline{\psi}, \overline{\Box\theta}, \overline{\downarrow\Box\chi}, \overline{\mathcal{F}\xi})$$

$$\downarrow\Box Q_1 y_n \cdots Q_n y_n \mathrm{BC}(\overline{\varphi}, \overline{@\psi}, \overline{\Box\theta}, \overline{\downarrow\Box\chi}, \overline{\mathcal{F}\xi}) \leftrightarrow \mathcal{F}Q_1 y_n \cdots Q_n y_n \mathrm{BC}(\overline{@\varphi}, \overline{\psi}, \overline{\Box\theta}, \overline{\downarrow\Box\chi}, \overline{\mathcal{F}\xi}).$$

Thus, in our original formula, we can replace any $\mathcal{F}$ with $\downarrow\Box$ or *vice versa*.   ■

To sum up, the following questions have yet to be answered about the relative **U**-expressive power of these languages:

- Is $\mathcal{L}^{1\mathrm{M}}(@, \downarrow) <_{\mathbf{U}} \mathcal{L}^{1\mathrm{M}}(\approx, @, \mathcal{F})$?
- Is $\mathcal{L}^{1\mathrm{M}}(@, \downarrow, \mathcal{F}) <_{\mathbf{U}} \mathcal{L}^{1\mathrm{M}}(\approx, @, \mathcal{F})$?
- Is $\mathcal{L}^{1\mathrm{M}}(@, \mathcal{F}) <_{\mathbf{U}} \mathcal{L}^{1\mathrm{M}}(@, \downarrow, \Pi)$ or $\mathcal{L}^{1\mathrm{M}}(@, \downarrow) <_{\mathbf{U}} \mathcal{L}^{1\mathrm{M}}(@, \mathcal{F}, \Pi)$?
- Is $\mathcal{L}^{1\mathrm{M}}(@, \downarrow, \mathcal{F}) <_{\mathbf{U}} \mathcal{L}^{1\mathrm{M}}(@, \downarrow, \Pi)$ or $\mathcal{L}^{1\mathrm{M}}(@, \downarrow, \mathcal{F}) <_{\mathbf{U}} \mathcal{L}^{1\mathrm{M}}(@, \mathcal{F}, \Pi)$?

We now finally turn to **UD**. Excluding E, $\approx$, and Π, the diagram in Figure 11 is still correct (again, the dashed arrows have not been determined). And again, Figure 9 is still correct when adding either E or $\approx$. But adding Π is even trickier than before, since we can no longer appeal to **Lemma D.10**. We still have the lack of inclusions mentioned above **Lemma D.12**. We also have the following lack of inclusions:

Lemma D.13 (*Inexpressibility of @ with Π*). $\mathcal{L}^{1\mathrm{M}}(\Pi) \not\leqslant_{\mathbf{UD}} \mathcal{L}^{1\mathrm{M}}(\approx, @)$.

Proof: Recall that $\mathcal{R}_1, w, w \leftrightarroweq_{\approx, @} \mathcal{R}_2, w, w$ (Figure 3). But the models are distinguished by $\exists x\, (\mathsf{Rich}(x) \land \Box\,(\mathsf{Poor}(x) \to \Sigma y\,(\neg\,\mathsf{Rich}(y) \land \neg\,\mathsf{Poor}(y))))$.   ■

Lemma D.14 (*Inexpressibility of @, Π with @, $\downarrow$, $\mathcal{F}$*). $\mathcal{L}^{1\mathrm{M}}(@, \Pi) \not\leqslant_{\mathbf{UD}} \mathcal{L}^{1\mathrm{M}}(\approx, @, \downarrow, \mathcal{F})$.

Proof: This immediately follows from **Proposition C.1**.   ■

However, we now have more inclusions. For example, $\mathcal{L}^{1\mathrm{M}}(\Pi) <_{\mathbf{UD}} \mathcal{L}^{1\mathrm{M}}(@, \downarrow)$ (just set $\Pi x \varphi \coloneqq \downarrow\Box\forall x\,@\varphi$).[37] Likewise, $\mathcal{L}^{1\mathrm{M}}(\Pi) <_{\mathbf{UD}} \mathcal{L}^{1\mathrm{M}}(@, \mathcal{F})$, though the proof is a bit more roundabout.[38] These inclusions are strict by **Lemma D.4**. The questions mentioned above for **U**-expressive power are still unanswered for **UD**-expressive power. And again, answering these questions suffices to settle the remaining inclusions.

---

[37]See, for instance, Hodes [1984b]; Forbes [1989]; Fine [2005]; Correia [2007].

[38]Here is the idea. First, every @ in φ is in the immediate scope of a $\mathcal{F}$, then $\Vdash_{\mathbf{U}} \Box\varphi \leftrightarrow \mathcal{F}@\varphi$. So every $\Box$ in a $\mathcal{L}^{1\mathrm{M}}(\Pi)$-formula can be replaced by $\mathcal{F}@$. Second, note that $\Vdash^{\mathrm{d}}_{\mathbf{UD}} \Pi x \varphi \leftrightarrow \Box\forall x\,@\varphi$. So having replaced every $\Box$ with $\mathcal{F}@$, every Π is in the scope of a $\mathcal{F}@$, so we can replace every Πx with $\Box\forall x\,@$. The result is a **UD**-equivalent $\mathcal{L}^{1\mathrm{M}}(@, \mathcal{F})$-formula.

REFERENCES

Areces, Carlos, Blackburn, Patrick, and Marx, Maarten. 1999. "Hybrid Logic is the Bounded Fragment of First-Order Logic." In *Proceedings of 6th Workshop on Logic, Language, Information and Computation*.

Areces, Carlos and ten Cate, Balder. 2007. "Hybrid Logics." In Patrick Blackburn, Frank Wolter, and Johan van Benthem (eds.), *Handbook of Modal Logic*, 821–868. Elsevier.

Bell, John Lane and Slomson, Alan Benjamin. 2006. *Models and Ultraproducts: An Introduction*. Courier Corporation.

Blackburn, Patrick, de Rijke, Maarten, and Venema, Yde. 2001. *Modal Logic*. Cambridge University Press.

Bricker, Phillip. 1989. "Quantified Modal Logic and the Plural De Re." *Midwest Studies in Philosophy* 14:372–394.

Chang, Chen Chung and Keisler, H Jerome. 1990. *Model Theory*. Third Edition. Courier Corporation.

Correia, Fabrice. 2007. "Modality, Quantification, and Many Vlach-Operators." *Journal of Philosophical Logic* 36:473–488.

Cresswell, Maxwell J. 1990. *Entities and Indices*. Kluwer Academic Publishers.

Crossley, John N and Humberstone, Lloyd. 1977. "The Logic of "Actually"." *Reports on Mathematical Logic* 8:11–29.

Davies, Martin and Humberstone, Lloyd. 1980. "Two Notions of Necessity." *Philosophical Studies* 38:1–30.

Fine, Kit. 1979. "Failures of the Interpolation Lemma in Quantified Modal Logic." *The Journal of Symbolic Logic* 44:201–206.

—. 1981. "Model Theory for Modal Logic - Part III Existence and Predication." *Journal of Philosophical Logic* 10:293–307.

—. 2005. "Prior on the Construction of Possible Worlds and Instants." In *Modality and Tense*, 133–175. Oxford University Press.

Forbes, Graeme. 1989. *Languages of Possibility*. Blackwell Publishers.

Fritz, Peter. 2012. "Modal Ontology and Generalized Quantifiers." *Journal of Philosophical Logic* 1–36.

Gabbay, Dov M. 1981. "Expressive Functional Completeness in Tense Logic." *Aspects of Philosophical Logic* 91–117.

Garson, James W. 2001. "Quantification in Modal Logic." In *Handbook of Philosophical Logic*, 267–323. Dordrecht: Springer Netherlands.

Gilbert, David R. 2015. "Actuality, Quantifiers and Actuality Quantifiers." *Logique et Analyse* .

Goranko, Valentin and Otto, Martin. 2007. "Model Theory of Modal Logic." *Handbook of Modal Logic* .

Hazen, Allen P. 1976. "Expressive Completeness in Modal Language." *Journal of Philosophical Logic* 5:25–46.

—. 1990. "Actuality and Quantification." *Notre Dame Journal of Formal Logic* 31:498–508.

Hodes, Harold T. 1984a. "Axioms for Actuality." *Journal of Philosophical Logic* 13:27–34.

—. 1984b. "On Modal Logics which Enrich First-order S5." *Journal of Philosophical Logic* 13:423–454.

—. 1984c. "Some Theorems on the Expressive Limitations of Modal Languages." *Journal of Philosophical Logic* 13:13–26.

Kocurek, Alexander W. 2015. "On the Expressivity of First-Order Modal Logic with "Actually"." In *Logic, Rationality, and Interaction*, 207–219. Berlin, Heidelberg: Springer Berlin Heidelberg.

—. 2016. "The Problem of Cross-world Predication." *Journal of Philosophical Logic* 1–46.

Lewis, David K. 1973. "Counterfactuals and Comparative Possibility." *Journal of Philosophical Logic* 2:418–446.

Needham, Paul. 1975. *Temporal Perspective: A Logical Analysis of Temporal Reference in English*. Ph.D. thesis, Upsala, Sweden.

Sider, Theodore. 2010. *Logic for Philosophy*. Oxford University Press.

Sturm, Holger and Wolter, Frank. 2001. "First-order Expressivity for S5-models: Modal vs. Two-sorted Languages." *Journal of Philosophical Logic* 30:571–591.

van Benthem, Johan. 1977. "Tense Logic and Standard Logic." *Logique et Analyse* 80:395–437.

—. 2010. "Frame Correspondences in Modal Predicate Logic." 1–14.

Vlach, Frank. 1973. *"Now" and "Then": A Formal Study in the Logic of Tense Anaphora*. Ph.D. thesis, UCLA.

Wehmeier, Kai F. 2001. "World Travelling and Mood Swings." In *Trends in Logic*, 257–260. Dordrecht: Springer Netherlands.

Yanovich, Igor. 2015. "Expressive Power of "Now" and "Then" Operators." *Journal of Logic, Language and Information* 24:65–93.