Kalpa Publications in Computing
Volume 18, 2024, Pages 163–178
LPAR 2024 Complementary Volume

A Case for Extensional Non-Wellfounded Metamodeling

Joseph Zalewski and Pascal Hitzler

Kansas State University
jzalewski@ksu.edu, phitzler@ksu.edu

Abstract

We introduce a notion of extensional metamodeling that can be used to extend knowledge representation languages, and show that this feature does not increase computational complexity of reasoning in many cases. We sketch the relation of our notion to various existing logics with metamodeling and to non-wellfounded sets, and discuss applications. We also comment on the usability of black-box reductions to develop reasoning algorithms for metamodeling.

1 Introduction

A common theme in logical languages for knowledge representation (KR) is *type separation*. For example, in description logics (DLs) [3] there are usually different types of syntactic entities for concepts/classes (such as 'Eagle') and individuals (such as a particular eagle). *Metamodeling* is the practice of using one type of entity in the way appropriate to a different type. For example, most knowledge-representation languages can express that an individual belongs to a class, but to say (something like) that a class belongs to a class is metamodeling. It is so called because ontologists (the W3C standard Web Ontology Language OWL is based on DLs [9]) often refer to an ontology as a 'model' of the individuals that populate it. To treat the classes and relations of the ontology as subjects of the same kind of modeling is 'meta-modeling'. [10] classifies the many applications of metamodeling into two kinds: domain-specific (e.g., in biology, a family may belong to an order), and general issues of the ontology-modeling process (e.g., associating a class with the researcher who added it to the ontology).

While working with reasoning over aggregate individuals, we realized we could do it elegantly by treating the class of the parts as an individual (see section 6). But this required an *'extensional'* notion of metamodeling, whereas most existing approaches are 'intensional', and for good reason (see below for explanation, and [10] for arguments in favor of the intensional; we have found that compelling use cases for the extensional are hard to find in the literature). In this paper, we demonstrate how intensional metamodeling can be simulated in an extensional-metamodeling language by careful ontology design, and argue that this is better than baking the somewhat mysterious concept of intension into the language semantics.

One line of semantics work that has taken extensionality seriously is that of Motz, Severi and Rohrer in e.g. [16] and [18], in which universes of interpretation contain some objects that just are sets of other objects, and an individual and a set can be the same in the most literal way. Proofs about these systems are quite complex, mainly because the structure of models is constrained by *well-foundedness* requirements (which flow naturally from their set-theoretic nature). We argue that this too is a feature best separated out from the logic. We propose an abstract semantics for metamodeling that is equivalent to a wellfounded-set semantics for certain ontologies designed to ensure this, but equivalent to a non-wellfounded-set semantics for other ontologies. Plausibly, it also allows ontologies to mix these assumptions, although the details of this require further work.

We show by a simple nondeterministic reduction that our metamodeling does not increase complexity of reasoning for many KR languages. This proof is 'black-box' in the sense that it uses an inclusive notion of 'logic' not referring to any particular syntax. As a point of interest, we also show that such a reduction cannot be made deterministic without additional postulates, unless $\mathsf{P} = \mathsf{NP}$. In most of the paper, we will focus on class-individual metamodeling, but in section 6 we show how a fundamental result generalizes to metamodeling with complex structures.

The paper is organized as follows: in Section 2, we define basic notions. In Section 3, we discuss a motivating example related to extensionality. In Section 4, we provide a motivating example related to wellfoundedness. In Section 5.1, we establish basic decidability and complexity facts about our metamodeling system, and analyze the 'optimality' of the proof. In Section 5.2, we connect our semantics to wellfounded and non-wellfounded set-theoretic semantics. In Section 5.3, we prove our simulation result for intensional metamodeling. In Section 6, we extend the main result of 5.1 to more complex metamodeling. In Section 7, we review related work.

2 Definitions

Assume there are fixed sets N_I of *individual names* and $N_{P,n}$ of *predicate names of arity* n for each positive natural number n, and these symbols have some binary encoding.

Let a *first-order structure*[1] $\mathcal{I}$ with respect to N_I, $N_{P,n}$ be a set $\Delta^{\mathcal{I}}$, called the *domain* of $\mathcal{I}$, together with a mapping $\cdot^{\mathcal{I}}_I : N_I \to \Delta^{\mathcal{I}}$, and mappings $\cdot^{\mathcal{I}}_{P,n} : N_{P,n} \to \mathcal{P}((\Delta^{\mathcal{I}})^n)$. (The domain models all the objects of a certain kind in some possible state of the universe, and the mappings model reference, of names to those objects and of predicates to relations among those objects.) We will write all the mappings as $\cdot^{\mathcal{I}}$ when no confusion is possible. We use A, B for class names (members of $N_{P,1}$) and a, b, c for individual names. A predicate of arity 2 (member of $N_{P,2}$) is a *role* in description logic terminology. Let $\mathcal{I}, \mathcal{J}$ be first-order structures, and $i : \Delta^{\mathcal{I}} \to \Delta^{\mathcal{J}}$. i is called an *isomorphism* if it is a bijection, and for all k, predicates $P \in N_{P,k}$, and $(x_1, ...x_k) \in (\Delta^{\mathcal{I}})^k$, $(x_1, ...x_k) \in P^{\mathcal{I}}$ iff $(i(x_1), ...i(x_k)) \in P^{\mathcal{J}}$, and if there exists such a mapping we say $\mathcal{I}, \mathcal{J}$ are isomorphic.

Let a *logic* L (with respect to N_I, $N_{P,n}$) be a set of strings, called *axioms* or *formulas* of L, containing symbols from N_I, $N_{P,n}$ and possibly other symbols, together with a relation $\models_L$ relating first-order structures over N_I, $N_{P,n}$ to axioms of L.[2] When the logic intended is clear, we will write $\models$ instead of $\models_L$. If S is a set of axioms, we write $\mathcal{I} \models S$ to mean $\mathcal{I} \models A$ for all $A \in S$. We write $S \models A$ ('S entails A') to mean that for every $\mathcal{I}$ such that $\mathcal{I} \models S$, $\mathcal{I} \models A$ also. A set S of axioms is *satisfiable* if there exists $\mathcal{I}$ such that $\mathcal{I} \models S$.

A logic will be called *isomorphism-invariant* if for any two isomorphic first-order structures $\mathcal{I}, \mathcal{J}$, and any axiom $\alpha \in L$, $\mathcal{I} \models_L \alpha$ iff $\mathcal{J} \models_L \alpha$. A logic will be called *decidable* if there is an algorithm which, given a finite set of axioms S, decides whether S is satisfiable.

We will sometimes use the notation $\mathcal{I} \models A \sqsubseteq B$ to mean $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$.

[1] Sometimes called an *interpretation*

[2] This relation may be a proper class relation.

Set Metamodeling Let L be a logic. We define an extended logic L^{M} as follows (see [18]):

Let $\approx$ be a new symbol not appearing in any axioms of L. Then the axioms of L^{M} are the axioms of L, plus all axioms of the forms $a \approx A$ and $a \not\approx A$, where a is an individual name and A is a class name (is in $N_{P,1}$). Let a *metamodeling interpretation* $\mathcal{I}$ be an interpretation $\mathcal{I}^*$ together with a relation $\approx^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}^*} \times \mathcal{P}(\Delta^{\mathcal{I}^*})$ which is a *local bijection*, i.e. for all $x, y \in \Delta^{\mathcal{I}}, X, Y \in \mathcal{P}(\Delta^{\mathcal{I}})$ such that $x \approx^{\mathcal{I}} X$ and $y \approx^{\mathcal{I}} Y$, $x = y$ iff $X = Y$. The relation $\models_{L^{\mathsf{M}}}$ between metamodeling interpretations and L^{M} formulas is defined as follows:

1. If A is an L axiom, $\mathcal{I} \models_{L^{\mathsf{M}}} A$ iff $\mathcal{I}^* \models_L A$
2. $\mathcal{I} \models_{L^{\mathsf{M}}} (b \approx B)$ iff $b^{\mathcal{I}^*} \approx^{\mathcal{I}} B^{\mathcal{I}^*}$
3. $\mathcal{I} \models_{L^{\mathsf{M}}} (b \not\approx B)$ iff it is not the case that $b^{\mathcal{I}^*} \approx^{\mathcal{I}} B^{\mathcal{I}^*}$

Define satisfiability etc. in the same way as for ordinary logics. From now on we will write, e.g., $b^{\mathcal{I}}$ instead of $b^{\mathcal{I}^*}$. (Technically, L^{M} is not a 'logic' in the sense just defined, since its $\models$ relation pertains to metamodeling interpretations, not just first-order structures. We may informally call it a 'logic' anyway, but when we refer to 'any logic L such that...', we mean something with semantics given by plain first-order structures.)

3 Example I: Intensionality

One of the debates in the metamodeling literature has been extensional vs. intensional semantics. In extensional metamodeling, to treat a class as an individual is to let that individual's name refer to the *extension* of the class, and in intensional semantics one lets the individual name refer to the *intension* of the class. This distinction is normally phrased in systems like Motik's [14]; we transfer it to our setting, where the syntax resembles [16]. Say that a logic satisfies Extensionality if the following inference is valid for all a, b, A, B:

$\{(a \approx A), (b \approx B), A \sqsubseteq B, B \sqsubseteq A\} \models_L (a \approx b)$

L^{M} satisfies Extensionality. To illustrate the contrast, we define another logic-forming operator:

Let L be a logic. Now let L^{IntM} be the logic L extended with the axioms $(a \approx A), (a \not\approx A)$, and define metamodeling interpretations as for L^{M}, except that we no longer require $\approx_{\mathcal{I}}$ to be injective (but it must still be functional): there may be two different $x, y \in \Delta^{\mathcal{I}}$ and one set X such that $x \approx_{\mathcal{I}} X$ and $y \approx_{\mathcal{I}} X$. This is obviously appropriate for a relation that means 'picking out a set' rather than 'being a set.' The difference between L^{IntM} and L^{M} parallels that of the *HiLog* and *injective HiLog* semantics discussed in [10]; that paper shows (in the appendix) that the latter is equivalent to the more usual *Henkin* semantics for extensional metamodeling.

It is not always desirable to have to choose one of these options. We use a classic example to illustrate. Suppose you are designing an ontology that includes the concept Raven, treated as a unary predicate (often called a 'concept name' in ontology modeling). You want to include some statements about properties of this concept. Let Black be the class of black objects, and $R' \equiv \mathsf{Black} \sqcap \mathsf{Raven}$. It will also turn out that $\mathsf{Raven} \equiv R'$, because $\mathsf{Raven} \sqsubseteq \mathsf{Black}$. But now suppose you want to say 'Raven is a biological concept'. We can do this with metamodeling: $(\mathsf{raven} \approx \mathsf{Raven})$, and $\mathsf{Biological}(\mathsf{raven})$. We might as well also treat the concept "black ravens" as an object that can have properties, by asserting $(r' \approx R')$. But we may want to assert that r' is *not* a biological concept, since we want $\mathsf{Biological}$ to refer only to concepts defined purely by biological features, and we don't consider color to be a biological feature. So $\neg\mathsf{Biological}(r')$.

With extensional metamodeling, this ontology is inconsistent, so it seems we must use intensional metamodeling. But now suppose we have another predicate of concepts, Instantiated. We intend this to hold of all concepts that apply to at least one object. (It applies to both Raven and R', in particular.) For this predicate, we would like to get the inference pattern: from $(a \approx A), (b \approx B), A \equiv B, \mathsf{Instantiated}(a)$, infer $\mathsf{Instantiated}(b)$. Extensional metamodeling would give us this inference power, by means of inferring $a = b$. But this looks like overkill, for something that might just be a quirk of the predicate Instantiated.

Note what peculiar feature of Instantiated makes this work: it is a predicate that depends only on the *extension* of a concept, whereas Biological is not. Instantiated is not a 'thick' property of concepts, but just a proxy for a property of another entity - the nonemptiness of the concept's extent. We can use these predicates together and get all the inferences we want, as follows:

Let hasExtent be a new binary predicate. Now to meta-model with the concept Raven, we introduce two individuals: raven and ravenExt, and assert hasExtent(raven,ravenExt), and (ravenExt $\approx$ Raven). Now we claim Biological(raven), but Nonempty(ravenExt). The predicate Instantiated can be recovered by the definition Instantiated $\equiv$ $\exists$hasExtent.Nonempty.[3]

Doing this has another advantage: it removes the connection between 'concept' objects and their extents from the purview of the logic and puts it in the hands of the ontology designer. Depending on the application, there may not be just one relation hasExtent, but many. The extent of Raven, in fact, changes over time. While a set is either Nonempty or not, a concept is Instantiated *at some time*. Many concepts have borderline cases (such as 'Forest'), and sometimes these are significant enough that the ontology designer does not want data that merely classifies things as 'a forest', but as a 'Forest according to authority or dataset X'. In that case the concept-object forest would have many extents, as well.

We hope we have argued well for this way of treating intensionality, but one thing we have not yet shown is that it is adequate to simulate intensional metamodeling built in to the semantics. We do this in section 5.3.

4 Example II: Wellfoundedness

'Wellfoundedness', in set theory, is the principle that there is no infinite descending chain of sets $S_1 \ni S_2 \ni S_3 \ni ....$ In particular, there cannot be a set that (directly or indirectly) contains itself. In the literal-equality semantics of [16], this has the consequence that, for instance, $\{(a \approx A), A(a)\}$ is always inconsistent.

Especially in the development of so-called *upper ontologies*, which are meant to be maximally general ontologies from which other ontologies can be created by refinement, an ontologist may want a class like 'Continuant,' which is supposed to subsume all objects, concrete or abstract, that persist through time. If the ontology also states properties *of* the concept Continuant itself, by use of a metamodeling axiom (continuant $\approx$ Continuant), then it will be appropriate to state, or at least not to contradict, the axiom Continuant(continuant), since a concept does persist through time. Our semantics does not render this inconsistent.

Nevertheless, consequences of wellfoundedness are often desirable, and can help detect bugs in an ontology. In an ontology for material objects, we might have some concept names referring to types of material objects, such as Metallic or Engine, and others referring to types of properties, such as Compositional or Structural. If we know Structural(engine) and Compositional(metallic), then wellfoundedness implies that the reverse containments cannot hold.

[3]Typical description logic syntax. In first-order syntax, $\forall x[\mathsf{Instantiated}(x) \leftrightarrow \exists y[\mathsf{hasExtent}(x, y) \wedge \mathsf{Nonempty}(y)]]$.

But the same functionality can be obtained with additional axioms. We might define 2 'level' classes, MaterialObject and Property-1, and declare $\forall x[\neg(\mathsf{MaterialObject}(x) \wedge \mathsf{Property\text{-}1}(x))]$, and Property-1(engine), Property-1(metallic), etc., and Metallic $\sqsubseteq$ MaterialObject, etc. As we had in the case of intensionality, implementing wellfoundedness at the level of ontology design provides flexibility: not only does it allow this ontology to be seamlessly integrated with an upper ontology using Continuant(continuant), it also allows us to finely control what can be a member of a class. Suppose we (using metamodeling) lump the known properties of Property-1's into a class called Property-2. Wellfoundedness implies that property-2 cannot be a Property-1. But it does not prevent, say, Property-1)t, for a particular toaster (a MaterialObject) t. Nor does it prevent Structural(t). Even with wellfoundedness, some domain-controlling axioms need to be written, to make a good hierarchy of meta-properties. So we argue that all decisions of this kind should be made at once, in ontology design, rather than leaving the developer to think through which axioms are consequences of semantic wellfoundedness and which are not. But also like intensionality, we would like to guarantee that this kind of maneuver can simulate all the effects of semantic wellfoundedness. We do this for a special case in section 5.2.

5 Results

5.1 Decidability and Complexity

Say that a logic L has *individual equality* (resp. *inequality)* if for every two individual names a, b there is an axiom X of L such that for all $\mathcal{I}$, $\mathcal{I} \models X$ iff $a^{\mathcal{I}} = b^{\mathcal{I}}$ (resp. $a^{\mathcal{I}} \neq b^{\mathcal{I}}$). We will write such an axiom "$(a = b)$" (resp. "$(a \neq b)$"), but this need not be the syntactic form of the axiom. We do, however, require that there is an algorithm to generate X given (the binary encoding of) a, b in polynomial time.

Likewise, a logic has *class equality (inequality)* if for every two class names A, B, there is an axiom $(A \equiv B)$ (respectively $(A \not\equiv B)$) such that for all $\mathcal{I}$, $\mathcal{I} \models (A \equiv B)$ iff $A^{\mathcal{I}} = B^{\mathcal{I}}$ ($\mathcal{I} \models (A \not\equiv B)$ iff $A^{\mathcal{I}} \neq B^{\mathcal{I}}$), and these axioms can be generated for each (A, B) in polynomial time. Our proof is somewhat similar to Motik's Theorem 6 in [14].

Theorem 1. *Let L be a decidable logic with individual and class equality and inequality axioms. Then L^{M}-satisfiability is decidable by an algorithm that runs in nondeterministic polynomial time and makes one call to an oracle for L-satisfiability.*

Proof. Let T be a finite set of L^{M}-axioms. An *identification* X for T is a set of axioms such that:

1. for every individual name a appearing in T and every class name A appearing in T, exactly one of $(a \approx A)$ or $(a \not\approx A)$ is in $X \cup T$

2. for every pair a, b of individual names appearing in T, exactly one of $a = b$ or $a \neq b$ is in X

3. for every pair A, B of class names appearing in T, exactly one of $A = B$ or $A \neq B$ is in X

4. if $(a \approx A), (b \approx B), (a = b) \in T$, then $A \equiv B \in T$

5. if $(a \approx A), (b \approx B), (A \equiv B) \in T$, then $a = b \in T$.

6. if $(a \approx A), (a = b), (A \equiv B) \in T$, then $(b \approx B) \in T$.

Let $L(X)$ denote the set of L-axioms in X.

Claim: T is L^{M}-satisfiable iff there exists an identification X for T such that $L(T) \cup L(X)$ is L-satisfiable.

Proof of claim: Suppose $\mathcal{I} \models_{L^{\mathsf{M}}} T$. Let X contain $(a \approx A)$ (where a, A appear in T) iff $a^{\mathcal{I}} \approx^{\mathcal{I}} A^{\mathcal{I}}$, and otherwise let X contain $(a \not\approx A)$. Also let X contain $a = b$ iff $a^{\mathcal{I}} = b^{\mathcal{I}}$, otherwise let X contain $a \neq b$. Likewise, let X contain $A \equiv B$ iff $A^{\mathcal{I}} = B^{\mathcal{I}}$, and $A \not\equiv B$ otherwise. Then X is an identification for T: the first three conditions are trivial, and the others follow because $\approx^{\mathcal{I}}$ is a local bijection. $\mathcal{I}^*$ (the first-order structure underlying $\mathcal{I}$) $\models_L L(T) \cup L(X)$.

Now suppose there exists an identification X for T such that $\mathcal{I} \models L(T) \cup L(X)$, for some first-order structure $\mathcal{I}$. Then we define a metamodeling structure $\mathcal{J}$ whose underlying first-order structure is $\mathcal{I}$, and define

$$\approx^{\mathcal{J}} = \{(y, Y) \mid \exists a, A[(a \approx A) \in X \wedge (a^{\mathcal{I}} = y) \wedge (A^{\mathcal{I}} = Y)]\}$$

$\approx^{\mathcal{J}}$ is a local bijection. For, let $x \approx^{\mathcal{J}} X$ and $y \approx^{\mathcal{J}} X$. Then there are a, b, A, B (not necessarily distinct) such that $x = a^{\mathcal{J}}, y = b^{\mathcal{J}}, X = A^{\mathcal{J}} = B^{\mathcal{J}}$ and $(a \approx A), (b \approx B) \in X$. But either $A \equiv B$ or $A \not\equiv B$ is in X, and the latter cannot hold, since $\mathcal{I} \models X$. Therefore $A \equiv B \in X$. Then by one of the postulates of identifications, $a = b \in T$, so $x = y$, as required. (The other direction of the proof of bijection is the same.) $\mathcal{J}$ also satisfies the axioms in X of the form $(a \approx A)$, by construction. Conversely, let $\mathcal{J} \models_{L^{\mathsf{M}}} (a \approx A)$. By construction there is some b, B such that $(b \approx B) \in X$, and $b^{\mathcal{J}} = a^{\mathcal{J}}$, $B^{\mathcal{J}} = A^{\mathcal{J}}$. Arguing as before, $a = b$ and $A \equiv B$ are in X. So by the last postulate of identifications, $(a \approx A) \in X$. So, $\mathcal{J}$ satisfies *exactly* the $(a \approx A)$ axioms that appear in X, and therefore exactly the $(a \not\approx A)$ axioms in X. This finishes the claim.

Now the algorithm, on input T, guesses an identification X, which is of size $O(|T|^2)$, and checks that $L(T) \cup L(X)$ is L-satisfiable. By the above claim, this checks L^{M} satisfiability of T. □

Corollary 2. *If L-satisfiability is in* NP *(resp.* PSPACE, EXP, NEXP, N2EXP*), and L has individual and class equality and inequality, then L^{M}-satisfiability is in* NP *(resp.* PSPACE, EXP, NEXP, N2EXP*).*

This corollary indicates that, by the rough metric of complexity class, most description logics do not become less tractable when we extend them with set metamodeling. However, the corollary does not mention *polynomial-time* logics, because a nondeterministic algorithm calling a P oracle does not suffice to put a problem in P. There is probably no way to improve the reduction in Theorem 1 to deterministic polynomial time without using additional properties of L, as we now show.

Theorem 3. *There exists a logic L with individual and class equality and inequality, such that L-satisfiability is in* P*, but L^{M}-satisfiability is* NP*-hard.*

Proof. Let B be some NP-complete language. So it has proofs of membership of length $n^{O(1)}$. Fix some reasonable polynomial-size encoding of partitions of $[1...n]$ and choose some efficient algorithm that, for each n, maps partitions of $[1...n]$ surjectively to strings of length $\leq n^{\Omega(1)}$. Thus we can use partitions of the set $[1...n^k]$, for a fixed k, as proofs of membership in B for instances of length n, instead of arbitrary strings. Now for each string x of length n, we choose a string of length $O(n^{O(1)})$ efficiently computable from x, which we call the 'axiom' ϕ_x, and assume x can be efficiently recovered from ϕ_x. For each n, let $a_1, ...a_{n^k}$ be some set of individual names efficiently enumerable given n, and likewise class names $A_1...A_{n^k}$. Any first-order structure $\mathcal{I}$

defines a partition of a_i by the equivalence relation $a_i \cong a_j$ iff $a_i^{\mathcal{I}} = a_j^{\mathcal{I}}$, and likewise A_i; we can also treat these as two partitions of the indices $i \leq n^k$. Let $K(x) : \{0,1\}^* \to \mathbb{N}$ be an injective function everywhere $> |x|^k$. Define satisfaction of ϕ_x as follows:

$\mathcal{I} \models_L \phi_x$ iff the two partitions of $[1...n^k]$ defined by $a_i, A_i, \mathcal{I}$ are equal, and are both partition proofs of membership for x in B, or the two partitions are different; and furthermore there are exactly $K(x)$ elements in $\Delta^{\mathcal{I}}$.

Let L be the logic of the axioms ϕ_n, and individual and class equality and inequality axioms with their expected semantics.

Claim 1: L-satisfiability is in P:

Given a set S of L-axioms, if there are two $\phi_x, \phi_{x'}$, $x \neq x'$, in it, we reject S as unsatisfiable, because a domain cannot contain exactly $K(x)$ and $K(x')$ elements. So we need only consider the case of a single ϕ_n and some equality and inequality axioms. Now in a domain of size $K(x) > n^k$, $\cdot^{\mathcal{I}}$ can clearly be chosen to induce any partition of a_i and any partition of A_i. It is also possible, given a set of equality and inequality axioms E on a_i, to decide in polynomial time whether there are 0, 1, or more than 1 partitions of $[1...n^k]$ induced on the a_i by structures satisfying E, and if there is exactly 1, return the unique such partition. The same, of course, holds for A_i. To decide satisfiability of S, do this for both a_i and A_i, and if either case returns '0 partitions', reject; otherwise, if either case returns '>1 partitions', accept - for there is some structure in which the two partitions are different, which therefore satisfies ϕ_n. Finally, if both cases return a unique partition, check if the partitions are equal. If not, accept; if so, check whether this partition is a proof of membership for x in B. If so, accept; otherwise reject.

Claim 2: L^{M}-satisfiability is NP-hard:

We reduce the NP-hard problem B. Given x, we decide whether $x \in B$ as follows: compute the axiom ϕ_n and check L-satisfiability of $S = \{\phi_x\} \cup \{(a_i \approx A_i \mid 1 \leq i \leq n^k)\}$. Any structure satisfying S must induce the same partition on both a_i and A_i because of the metamodeling axioms, but it also satisfies ϕ_x, so this partition is a proof of membership for x in B. Conversely, if there is a partition proof P of membership for x in B, there exists a structure $\mathcal{I}$ of $K(x)$ elements inducing this partition on both a_i, A_i; now define the metamodeling relation $\approx_{\mathcal{I}}$ such that $a_i^{\mathcal{I}} \approx_{\mathcal{I}} A_j^{\mathcal{I}}$ iff i, j are in the same block in P. It is easy to see this is a local bijection, and it satisfies the metamodeling axioms in S. Thus, S is L-satisfiable iff x has some partition proof of membership, iff x is in B. □

Nevertheless, this complexity blowup seems to be pathological, and efficient deterministic reductions are possible when certain stronger assumptions hold. One case in which metamodeling does not increase complexity of reasoning problems in P is for entailment problems in logics with so-called *consequential models*, which include most of the logics of the $\mathcal{EL}$ family [18].[4] Note that the results in [18] are for a stricter semantics than ours (wellfoundedness). One of the logics in the $\mathcal{EL}$ family provides the semantics for OWL EL, an OWL language profile used in many influential ontologies, including SNOMED-CT [4] and the Gene Ontology [2][1].

5.2 Significance of the Semantics

The set metamodeling semantics we have presented in this paper is not obviously correct. It seems obvious that 'identity of a class with an entity' can be approximated by a binary relation between classes and entities, and if this relation is to resemble identity, it should be, among other things, a local bijection. But among what other things? It is not apparent how much

[4] This seems to be an unpublished manuscript

is lost by replacing identity with a relation. The results of this section justify the use of our semantics in some situations.

For any logic L, we defined the logics L^{M} and L^{IntM} using our metamodeling-relation semantics; now define the logic L^{TM} similarly, by extending L with the axioms $(a \approx A)$ and $(a \not\approx A)$, and extending $\models_L$ with the conditions:

1. $\mathcal{I} \models_{L^{\mathsf{TM}}} (a \approx A)$ iff $a^{\mathcal{I}} = A^{\mathcal{I}}$
2. $\mathcal{I} \models_{L^{\mathsf{TM}}} (a \not\approx A)$ iff $a^{\mathcal{I}} \neq A^{\mathcal{I}}$

L^{TM}, L^{IntM} and L^{M} have exactly the same formulas, but different semantics. L^{TM} has the most literal possible semantics for metamodeling - $(a \approx A)$ means that the (object) name a and the (class) name A actually refer to the same thing, which is an object that happens to be a class. Notice that nothing in the definition of first-order structures prevents an element of a structure from also being a set of elements in the same structure, but there is something weird about referring to a situation like this in a semantic clause.[5] Normally elements of a structure are thought of as featureless, and the only information about them relevant to the truth of axioms is provided by their relation to the mapping $\cdot^{\mathcal{I}}$. Nevertheless, since any other condition would leave open the same objection as the semantics of L^{M} (that we are not taking metamodeling seriously enough) we will use L^{TM} as a baseline for comparison.

Cyclicity The following theorem and lemma are true only when we switch our metamathematical perspective from standard ZFC set theory to Boffa's *non-wellfounded set theory* BAFA (see [17][13]), a set theory in which the wellfoundedness principle fails, and furthermore there are many non-wellfounded sets. However, by doing this we get a very good result: L^{M}- and L^{TM}-satisfiability hold of exactly the same sets of axioms. Boffa's set theory has been criticized (see [17]) as too 'intensional', which would make its use contrary to our aims; however, it also has its defenders (see [13]).

Let a *decoration* d of a directed graph G be a mapping from nodes in $V(G)$ to sets such that $d(x) = \{d(y) \mid (x, y) \in E(G)\}$. Let a pointed graph be a graph with a distinguished node, and call a pointed graph *accessible* if every node is reachable from the distinguished node. A graph is *extensional* if there are no two distinct nodes with the same successors. Boffa's antifoundation axiom says that every accessible extensional pointed graph has an injective decoration. The pointing and accessibility condition can mostly be ignored, as we see:

Lemma 4 (in BAFA)**.** *For any extensional graph G, there is an injective decoration.*

Proof. If G, x is accessible for some choice of distinguished $x \in V(G)$, then we are done. Suppose there is no such x. Let u be an object not in $V(G)$, and consider the graph G' with nodes $V \cup \{u\}$, whose edges are $E(G) \cup \{(u, g) \mid g \in V(G)\}$. Then (G', u) is an accessible pointed graph. It is also extensional: u is a successor of nothing, so it suffices to show no node in $V(G)$ has the same set of successors as u, namely, all of $V(G)$. But this is the case, since if there were such a node G would be accessible. Thus by Boffa's axiom there is an injective decoration d of G'. Further, $d(u)$ is not in any $d(g), g \in V(G)$, since u is not a successor of any g. So the restriction of d to G is also a decoration of G. □

Call a graph *ur-extensional* if no two distinct nonterminal nodes (i.e., nodes having some successor) have the same set of successors. Let G be an ur-extensional graph and N a set of

[5] Motz et al. [16] make it less strange by prescribing that models must be built up from $\mathbb{N}$ by powersets, but we do not strictly need to do this for $\models_{L^{\mathsf{TM}}}$ to make sense, so to make our comparison simpler, we don't.

nodes with up to one terminal node in it. An N-*ur-decoration* of G is a mapping d from nodes to sets such that for each node $x \in N$, $d(x) = \{d(y) \mid (x, y) \in E(G)\}$. The idea of this is that it treats the chosen nonterminal nodes as pictures of nonempty sets, and up to one terminal node as a picture of the empty set, and all other terminal nodes as 'urelements', which have internal structure we don't care about except that they are unique.

Lemma 5 (in BAFA)**.** *For every ur-extensional graph G and set $N \subseteq V(G)$ containing up to one terminal node, G has an injective N-ur-decoration, such that when terminal $x \notin N$, $d(x) \cap d[V(G)] = \emptyset$.*

Proof. Let G be ur-extensional. Let κ be the cardinality of the nodes of G. If necessary, replace the nodes of G so that none of them are ordinals $< \kappa + 2$. For each ordinal $\alpha < \kappa + 2$, add to G a node z_α, and an edge (z_β, z_α) whenever $\alpha < \beta$. Number the nodes of G as g_α for α from 1 to $\kappa + 1$, and add to G the edge (g_α, z_α) whenever g_α is terminal in G and not in N. Also add the edge (z_α, z_α) for all α. Call the result G'. G' is extensional:

Choose any two distinct nodes, and we can show they have different successors. Case I: g_α and g_β, both nonterminal. Different since G is ur-extensional. Case II: g_α in $V(G)$, g_β terminal, not in N: Then g_β has successor z_β which g_α does not. Case III: g_α in $V(G)$, g_β terminal in N: g_α has some successor and g_β does not. Case IV: g_α, z_β: g_α does not have successor z_0 (since we started numbering g_α at 1), but z_α does. Case V: z_α, z_β: by wellfoundedness of ordinals, (wlog) $\alpha < \beta$. Then z_β is a successor of z_β but not of z_α.

Thus, there is an injective decoration d of G' by lemma 4. d is also an ur-decoration of G: let x be nonterminal in G. Then its successors in G' are its successors in G, so $d(x) = \{d(y) \mid (x, y) \in E(G')\} = \{d(y) \mid (x, y) \in E(G)\}$ as required.

The last condition of the lemma follows since d is injective, and points not in N have only z's as successors. □

We prove results in this section for possibly-infinite sets of axioms partly just to show it can be done, but also to facilitate future applications in which a logic might employ axioms equivalent to an infinite set of basic metamodeling axioms. Finiteness plays no part in the possibility of interpreting our semantics set-theoretically.

Theorem 6 (in BAFA)**.** *For any isomorphism-invariant logic L, and any set of L^{M} axioms S (which is to say a set of L^{TM} axioms), S is satisfiable in L^{M} iff S is satisfiable in L^{TM}.*

Proof. If: Let $\mathcal{I} \models_{L^{\mathsf{TM}}} S$. We augment $\mathcal{I}$ with the relation $\approx^{\mathcal{I}}$ defined, for $x \in \Delta^{\mathcal{I}}$, $X \subseteq \Delta^{\mathcal{I}}$, by ($x \approx^{\mathcal{I}} X$ iff $x = X$). This is a local bijection, since it is just a subrelation of the identity relation, so $\mathcal{I}, \approx^{\mathcal{I}}$ is a metamodeling interpretation. Now for any L axiom $B \in S$, $\mathcal{I} \models_L B$, so $\mathcal{I}, \approx_{\mathcal{I}} \models_{L^{\mathsf{M}}} B$. And if $(a \approx A) \in S$, then $\mathcal{I} \models_{L^{\mathsf{TM}}} (a \approx A)$, so $a^{\mathcal{I}} = A^{\mathcal{I}}$, so $a^{\mathcal{I}} \approx^{\mathcal{I}} A^{\mathcal{I}}$, so $\mathcal{I} \models_{L^{\mathsf{M}}} (a \approx A)$. The same for $(a \not\approx A)$. Thus S is L^{M} satisfiable.

Only if: Let $\mathcal{I}, \approx^{\mathcal{I}}$ be some model of S in L^{M}. Define the relation of 'pseudomembership' $\in^P$: $x \in^P y$ iff $y \approx^{\mathcal{I}} Y$ for some set Y and $x \in Y$. Then $\Delta^{\mathcal{I}}, \ni^P$ is a directed graph. We claim it is ur-extensional: suppose not, and let x, y be distinct nonterminal points with the same set P of $\ni^P$ successors ('pseudo-elements'). Then $x \approx X$ and $y \approx Y$ for some (unique) sets (if not, they are $\ni^P$-terminal), and by definition of pseudo-member, $X = P = Y$. Thus by local bijectivity of $\approx$, $x = y$, a contradiction.

Let N be the set of $\ni^P$-nonterminal nodes in $\Delta^{\mathcal{I}}$, plus the node x such that $x \approx^{\mathcal{I}} \emptyset$, if there is one. There is an injective N-ur-decoration d of $\Delta^{\mathcal{I}}, \ni^P$ as described in Lemma 5. Let $\Delta^{\mathcal{J}} = d[\Delta^{\mathcal{I}}]$, and let $P^{\mathcal{J}} = \{(d(x_1), ...d(x_n)) \mid (x_1, ...x_n) \in P^{\mathcal{I}}\}$ for each predicate and $a^{\mathcal{J}} = d(a^{\mathcal{I}})$ for each individual. d is an isomorphism between $\mathcal{I}$ and $\mathcal{J}$, so $\mathcal{J} \models_L$ all L-axioms in S. Let $(a \approx A) \in S$. Then $a^{\mathcal{I}}$ is in N, so $a^{\mathcal{J}} = d(a^{\mathcal{I}}) = \{d(y) \mid (a^{\mathcal{I}} \ni^P y)\} = \{d(y) \mid (y \in A^{\mathcal{I}})\} = d[A^{\mathcal{I}}] = A^{\mathcal{J}}$. Let $(a \not\approx A) \in S$. There are two cases: Case I: $a^{\mathcal{I}} \approx X$ for some set X: so $X \neq A^{\mathcal{I}}$, but $a^{\mathcal{I}} \in N$, so $a^{\mathcal{J}} = d[X] \neq d[A^{\mathcal{I}}] = A^{\mathcal{J}}$, by injectivity of $d[\cdot]$. Case II: $a^{\mathcal{I}} \notin N$. Then by the last clause of lemma 5, $a^{\mathcal{J}}$ is disjoint from $\Delta^{\mathcal{J}}$ but not empty. Thus it is not a subset of $\Delta^{\mathcal{J}}$ and cannot be $A^{\mathcal{J}}$. □

This theorem shows that any reasoners developed for L^{M} can be thought of, if you prefer, as reasoners for L^{TM}, under the condition that you accept non-wellfounded set theory. We also sketch the connection between L^{M} and L^{TM} under ordinary metamathematics.

Acyclicity Call a set of L^{M} formulas S *(semantically) acyclic* if there is a wellfounded quasiordering $\leq$ on the individual names appearing in S, such that if $(a \approx A) \in S$, $\mathcal{I}, \approx_{\mathcal{I}} \models_{L^{\mathsf{M}}} S$ and $b^{\mathcal{I}} \in A^{\mathcal{I}}$, then $b < a$.

Lemma 7. *Let S be a set of L^{M} axioms. If S is satisfiable, it is satisfied by a metamodeling interpretation $\mathcal{I}$ in which $x \approx_{\mathcal{I}} X$ only holds for x, X where there is a, A such that $x = a^{\mathcal{I}}, X = A^{\mathcal{I}}, (a \approx A) \in S$. Furthermore, if there is an interpretation satisfying exactly the axioms S, there is an interpretation with the above feature, satisfying exactly the axioms S.*

Proof. Let $\mathcal{I}, \approx_{\mathcal{I}} \models S$. Now consider $\approx'$, the relation which contains just the pairs $a^{\mathcal{I}}, A^{\mathcal{I}}$ such that $(a \approx_{\mathcal{I}} A) \in S$. By definition of satisfaction of these kind of axioms, $\approx' \subseteq \approx_{\mathcal{I}}$. Therefore $\approx'$ is a local bijection. But $\mathcal{I}, \approx' \models (a \approx A)$ for all $(a \approx A) \in S$, so $\mathcal{I}, \approx' \models_{L^{\mathsf{M}}} S$. Clearly if $\mathcal{I}, \approx^{\mathcal{I}} \models (a \not\approx A)$, then so does $\mathcal{I}, \approx'$. □

Theorem 8. *Let S be semantically acyclic, and L be isomorphism-invariant. Then S is satisfiable in L^{M} iff S is satisfiable in L^{TM}.*

Proof. If: Same as the if-direction of Theorem 6.

Only if: Let $\mathcal{I}, \approx_{\mathcal{I}}$ be a L^{M} model of S. Assume without loss of generality that no element of $\Delta^{\mathcal{I}}$ is a set indirectly containing another member of $\Delta^{\mathcal{I}}$, and that $\approx_{\mathcal{I}}$ holds only on pairs $a^{\mathcal{I}}, A^{\mathcal{I}}$ where $(a \approx A) \in S$, which we can do by lemma 7. Divide the individual names in S into disjoint 'rank' sets R_α indexed by ordinals, starting at 1, such that if $a < b$, $a \in R_\alpha$, $b \in R_\beta$, then $\alpha < \beta$. This can be done by (transfinite) induction. Let the rank $\mathsf{rank}(a)$ of a name be the unique ordinal α such that $a \in R_\alpha$. Let the *rank* $\mathsf{rank}(x)$ of an element $x \in \Delta^{\mathcal{I}}$ be 0 if there is no set X where $x \approx^{\mathcal{I}} X$, and otherwise the least rank among names a where $a^{\mathcal{I}} = x$, and $(a \approx A) \in S$ for some A. Such a exists, by our assumption from lemma 7. For each α, define a partial mapping J_α on elements of rank $\leq \alpha$ by induction, as follows:

J_0 is the identity map.

$J_{\alpha+1}$ is equal to J_α on elements of rank $\leq \alpha$, and for $a^{\mathcal{I}}$ of rank $\alpha + 1$, $J_{\alpha+1}(a^{\mathcal{I}}) = J_\alpha[X]$, where X is the unique set such that $a^{\mathcal{I}} \approx_{\mathcal{I}} X$.

J_β, for β a limit ordinal, is equal to J_α on elements of rank α for all $\alpha < \beta$. For $a^{\mathcal{I}}$ of rank β, $J_\beta(a^{\mathcal{I}}) = [\bigcup_{\alpha<\beta} J_\alpha][X]$, where X is the unique set such that $a^{\mathcal{I}} \approx_{\mathcal{I}} X$.

J_α is well-defined. In the inductive cases, X exists because, by hypothesis, $a^{\mathcal{I}}$ is not of rank 0. Assume without loss of generality that a is chosen so $\mathsf{rank}(a) = \mathsf{rank}(a^{\mathcal{I}})$, and $(a \approx A) \in S$ (a gives x its rank). We should prove X is a subset of the domain of J_α (resp. the combined domains of J_α): $X = A^{\mathcal{I}}$ for some A where $(a \approx A) \in S$, by assumption. Now let $y \in A^{\mathcal{I}}$. If $\mathsf{rank}(y) > 0$, then $y \approx Y$ for some set Y, and there is b, B where $b^{\mathcal{I}} = y, B^{\mathcal{I}} = Y, (b \approx B) \in S$, and $\mathsf{rank}(b) = \mathsf{rank}(y)$. So by acyclicity, $b < a$, so $\mathsf{rank}(b) < \mathsf{rank}(a) = \alpha+1$ (resp. β). Otherwise $\mathsf{rank}(y) = 0$. This proves that all elements of X have rank less than that of $a^{\mathcal{I}}$, as required.

Each J_α is injective, by induction on α:

For $\alpha = 0$, trivial.

For level $\gamma = \alpha+1$, or γ a limit ordinal: first we prove $J_\gamma(a^{\mathcal{I}})$ is not in the range of J_α for any $\alpha < \gamma$. Suppose it is, and let x be an element of rank $\alpha < \gamma$ such that $J_\alpha(x) = J_\gamma(a^{\mathcal{I}})$. Since this is a set indirectly containing elements of $\Delta^{\mathcal{I}}$, we know α is not 0. Now, by construction, $J_\gamma(a^{\mathcal{I}}) = (\cup_\beta J_\beta)[Y]$ for some set Y such that $x \approx_{\mathcal{I}} Y$ and some set of ordinals $\beta < \gamma$, where Y is contained in the domain of $(\cup_\beta J_\beta)$. But $(\cup_\beta J_\beta)[Y] = J_\gamma[Y]$, and J_β are a nested sequence of injective functions, by inductive hypothesis. Thus their union is injective and a subfunction of J_γ, and we get $X = Y$. Then since $\approx_{\mathcal{I}}$ is a local bijection, $a^{\mathcal{I}} = x \in X$, which makes the rank of $a^{\mathcal{I}}$ less than $\mathsf{rank}(a^{\mathcal{I}})$, a contradiction.

Thus it suffices to show that for any two distinct x, y of rank γ exactly, $J_\gamma(x) \neq J_\gamma(y)$, but this follows from the construction since J_α is injective for $\alpha < \gamma$, and the sets X_x, X_y such that $x \approx_{\mathcal{I}} X_x$ and $y \approx_{\mathcal{I}} X_y$ are distinct.

Now take the union of all the J_α; call it J. J is injective. Let $\Delta^{\mathcal{J}}$ be the image $J[\Delta^{\mathcal{I}}]$, and define $\cdot^{\mathcal{J}}$ as in theorem 6, so that J is an isomorphism of $\mathcal{I}$ to $\mathcal{J}$. Now for each axiom $(a \approx A) \in S$, we have $a^{\mathcal{I}} \approx^{\mathcal{I}} A^{\mathcal{I}}$, so $J(a^{\mathcal{I}}) = J[A^{\mathcal{I}}]$, so $a^{\mathcal{J}} = A^{\mathcal{J}}$. Likewise, let $(a \not\approx A) \in S$. Case I: $a^{\mathcal{I}}$ is not $\approx^{\mathcal{I}} X$ for any X. Then $a^{\mathcal{I}}$ has rank 0, and $a^{\mathcal{J}} = a^{\mathcal{I}}$. But $A^{\mathcal{J}}$ is a set indirectly containing members of $\Delta^{\mathcal{I}}$, so it cannot be $a^{\mathcal{I}}$. Case II: $a^{\mathcal{I}} \approx^{\mathcal{I}} B^{\mathcal{I}}$ for some $B^{\mathcal{I}} \neq A^{\mathcal{I}}$. Then $a^{\mathcal{J}} = B^{\mathcal{J}} = J[B^{\mathcal{I}}] \neq J[A^{\mathcal{I}}] = A^{\mathcal{J}}$, because J is injective.

Thus $\mathcal{J} \models_{L^{\mathsf{TM}}} S$.

□

Practical Consequences The designer of an ontology cannot do anything with theorem 8 unless he can make the ontology semantically acyclic. Fortunately, in description logics extending $\mathcal{ALCO}$ (see e.g. [7] for this logic), acyclicity can be enforced by a syntactic condition. The following definition uses some $\mathcal{ALCO}$ syntax:

Suppose there is a wellfounded quasiordering $<$ on individual names in S. Let S be called *(syntactically) acyclic* if, for all pairs (a, b) of individual names appearing in S, where $b \not< a$, and axioms $(a \approx A) \in S$, we have $(\{b\} \sqcap A \sqsubseteq \bot) \in S$.

Proposition 9. *If S is syntactically acyclic, S is semantically acyclic.*

Proof. Trivial □

Here is our point of departure from the line of work of [16] [18]. In those systems, wellfoundedness is assumed in the logic itself; ontologies that imply a cyclic containment of sets are reported as inconsistent by a complete reasoner. In our approach, this does not happen - the ontology designer must ensure acyclicity by asserting it. Theorem 8 shows that if this is done ‘all the way’, reasoning becomes equivalent to reasoning with a wellfounded-set-based semantics like [16], but it is up to the ontologist to decide when and how to enforce acyclicity.

5.3 Simulation of Intensional Semantics by Extensional

To describe the logics for which the following simulation works, we introduce another generic operation. For any first-order structure $\mathcal{I}$, and set $S \subseteq \Delta^{\mathcal{I}}$, let the S-*reduct* $\mathcal{J}$ of $\mathcal{I}$ be the structure with domain S, and where for any predicate $P \in N_{P,k}$, $P^{\mathcal{J}} = P^{\mathcal{I}} \cap S^k$.[6]

[6] If $a^{\mathcal{I}} \notin U^{\mathcal{I}}$ for some individuals a, we can map them into $U^{\mathcal{I}}$ arbitrarily.

For a set of L axioms S, and a class name U, let a U-*localization* of S be a set of axioms S' such that for all $\mathcal{I}$, if $\mathcal{I} \models_L S'$ and $\mathcal{I} \models_L A \sqsubseteq U$ for all class names A appearing in S and $U(a)$ for all individual names a appearing in S, then the $U^{\mathcal{I}}$-reduct of $\mathcal{I}$ satisfies S; and if the $U^{\mathcal{I}}$-reduct of $\mathcal{I}$ satisfies S, then $\mathcal{I} \models S'$. Say that a logic L admits localization if for all sets of L-axioms S and class names U not appearing in S, there is a U-localization of S.

Let L be a logic, and suppose that for any role name $r \in N_{P,2}$ and individual names a, b, L has an axiom $r(a, b)$ such that $\mathcal{I} \models_L r(a, b)$ iff $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ (that is, L can express role assertions). Suppose L can also assert functionality of roles: that is, there is an axiom $F(r)$ such that $\mathcal{I} \models_L F(r)$ iff $r^{\mathcal{I}}$ is a functional relation. Let S be a set of axioms in the logic L^{IntM}. Let $r \in N_{P,2}$ be a role name not appearing in S, and for each $a \in N_I$ appearing in S, let a' be a new individual name not appearing in S. Define the set S' as follows: Assume without loss of generality that every class name appearing in S appears in an L axiom in S. Let $L(S)$ be the part of S consisting of L axioms and $M(S)$ the part consisting of metamodeling axioms. Let U be a class name not appearing in S, and let S'' be a U-localization of $L(S)$. Let S' be S'' extended with the axiom $r(a, a')$ for each a, and $F(r)$. Let π be a mapping that replaces $(a \approx A)$ with $(a' \approx A)$ and $(a \not\approx A)$ with $(a' \not\approx A)$, and on L-axioms $\alpha \in S$, let $\pi(\alpha)$ be a set of L axioms that is a U-localization of $\{\alpha\}$.

Theorem 10. *For any axiom α of L^{IntM}, $S \models_{L^{\mathsf{IntM}}} \alpha$ iff $S' \cup \pi[M(S)] \models_{L^{\mathsf{M}}} \pi(\alpha)$.*

Proof. Suppose $S \models_{L^{\mathsf{IntM}}} \alpha$, and $\mathcal{I} \models_{L^{\mathsf{M}}} S' \cup \pi[M(S)]$. Assume without loss of generality that $\approx_{\mathcal{I}}$ only holds between named entities (Lemma 7). Take the reduct of $\mathcal{I}$ to $U^{\mathcal{I}}$, and turn it into a metamodeling interpretation $\mathcal{I}'$ by defining $x \approx_{\mathcal{I}'} X$ iff for some a, $x = a^{\mathcal{I}}$ and $(a')^{\mathcal{I}} \approx_{\mathcal{I}} X$. $\mathcal{I}'$ is an L^{IntM} metamodeling interpretation, i.e., $\approx_{\mathcal{I}'}$ is functional, because $r^{\mathcal{I}}$ relates each $a^{\mathcal{I}}$ to $(a')^{\mathcal{I}}$, and $r^{\mathcal{I}}$ is functional. So $(a')^{\mathcal{I}}$ is determined uniquely by x. Furthermore, $\mathcal{I}' \models_{L^{\mathsf{IntM}}} S$: $L(S)$ holds in $\mathcal{I}'$ because the localization S'' holds in $\mathcal{I}$. So let $(a \approx A)$ be an axiom in $M(S)$. $\mathcal{I} \models_{L^{\mathsf{M}}} (a' \approx A)$, so $(a')^{\mathcal{I}} \approx_{\mathcal{I}} A^{\mathcal{I}}$, so by definition, $a^{\mathcal{I}'} = a^{\mathcal{I}} \approx_{\mathcal{I}'} A^{\mathcal{I}} = A^{\mathcal{I}'}$. (Equality of these sets follows since $\mathcal{I} \models A \sqsubseteq U$.) Let $(a \not\approx A)$ be in $M(S)$. Then similarly, $(a')^{\mathcal{I}} \not\approx_{\mathcal{I}} A^{\mathcal{I}}$, and this equivalence is necessary for $a^{\mathcal{I}} \approx_{\mathcal{I}'} A^{\mathcal{I}}$, since $(a')^{\mathcal{I}}$ is functionally determined by $a^{\mathcal{I}}$. So $a^{\mathcal{I}'} = a^{\mathcal{I}} \not\approx_{\mathcal{I}'} A^{\mathcal{I}} = A^{\mathcal{I}'}$.

Since $\mathcal{I}' \models_{L^{\mathsf{IntM}}} S$, also $\mathcal{I}' \models_{L^{\mathsf{IntM}}} \alpha$. So $\mathcal{I} \models_{L^{\mathsf{M}}} \pi(\alpha)$: by a similar argument to the above, if α is a metamodeling axiom, and otherwise, by definition of localization.

Converse: Let $S' \cup \pi[M(S)] \models_{L^{\mathsf{M}}} \pi(\alpha)$, and let $\mathcal{I} \models_{L^{\mathsf{IntM}}} S$. For each set X such that some $a^{\mathcal{I}}$ is related to X by $\approx_{\mathcal{I}}$, extend the domain of $\mathcal{I}$ with a new object $e[X]$. Let X_a denote the unique set related to $a^{\mathcal{I}}$. For each $a^{\mathcal{I}}$ not related to any X, extend the domain with a new object $e[a^{\mathcal{I}}]$. Define $\mathcal{I}'$ with this extended domain, such that $P^{\mathcal{I}'} = P^{\mathcal{I}}$ for predicates appearing in S, and $U^{\mathcal{I}'} = \Delta^{\mathcal{I}}$, and $\approx_{\mathcal{I}'} = \{(e[X_a], X_a)\}$. Let $(a')^{\mathcal{I}'} = e[X_a]$ if X_a exists, otherwise $e[a^{\mathcal{I}}]$. Let $r^{\mathcal{I}'} = \{(a^{\mathcal{I}}, (a')^{\mathcal{I}})\}$. $\mathcal{I}'$ is an L^{M} metamodeling interpretation, because by construction $\approx_{\mathcal{I}'}$ is a local bijection. Now $\mathcal{I}' \models_{L^{\mathsf{M}}} S'$: $F(r)$ and $r(a, a')$ hold by definition of $r^{\mathcal{I}'}$. $\mathcal{I}$ is the U-reduct of $\mathcal{I}'$, so $\mathcal{I}' \models_L S''$ by definition of localization. And $\mathcal{I}' \models_{L^{\mathsf{M}}} \pi[M(S)]$: for any $(a' \approx A) \in \pi[M(S)]$, $(a \approx A) \in M(S)$, so $\mathcal{I} \models_{L^{\mathsf{IntM}}} (a \approx A)$, so $a^{\mathcal{I}} \approx_{\mathcal{I}} A^{\mathcal{I}}$, so $X_a = A^{\mathcal{I}}$, so $e[X_a] \approx_{\mathcal{I}'} A^{\mathcal{I}}$, as needed. For any $(a' \not\approx A) \in \pi[M(S)]$, $(a \not\approx A) \in M(S)$, so $a^{\mathcal{I}} \not\approx_{\mathcal{I}} A^{\mathcal{I}}$, so $X_a \neq A^{\mathcal{I}}$ (or X_a fails to exist), therefore $(a')^{\mathcal{I}'} \not\approx_{\mathcal{I}'} A^{\mathcal{I}}$, so $\mathcal{I}' \models_{L^{\mathsf{IntM}}} (a' \not\approx A)$. This finishes the claim that $\mathcal{I}' \models_{L^{\mathsf{M}}} S' \cup \pi[M(S)]$.

Therefore $\mathcal{I}' \models_{L^{\mathsf{M}}} \pi(\alpha)$. But then $\mathcal{I} \models_{L^{\mathsf{IntM}}} \alpha$. This is by definition of localization if $\alpha \in L$. Otherwise: let $\mathcal{I}' \models_{L^{\mathsf{M}}} (a' \approx A)$. Then $(a')^{\mathcal{I}'} = e[X_b]$ and $A^{\mathcal{I}} = A^{\mathcal{I}'} = X_b$, for some b. But $(a')^{\mathcal{I}'}$ has this form only if $a^{\mathcal{I}} \approx_{\mathcal{I}} X_a$. Therefore $\mathcal{I} \models_{L^{\mathsf{IntM}}} (a \approx A)$. Finally, let $\mathcal{I} \models_{L^{\mathsf{M}}} (a' \not\approx A)$. Then either $(a')^{\mathcal{I}'} = e[X_b]$ for some b, or $(a')^{\mathcal{I}'} = e[a^{\mathcal{I}}]$. In case 1, if $a^{\mathcal{I}} \approx_{\mathcal{I}} A^{\mathcal{I}}$, then $X_a = A^{\mathcal{I}}$, but since $(a')^{\mathcal{I}'} = e[X_b]$, we have $X_b = X_a$. But X_b cannot be $A^{\mathcal{I}}$, since $\mathcal{I}' \models_{L^{\mathsf{M}}} (a' \not\approx A)$. This is a contradiction. In case 2, $a^{\mathcal{I}}$ is related to nothing by $\approx_{\mathcal{I}}$, so $\mathcal{I}' \models_{L^{\mathsf{IntM}}} (a \not\approx A)$.

Thus $\mathcal{I}' \models_{L^{\mathsf{IntM}}} \alpha$ in all cases.

□

This translation has a natural meaning: we expand our ontology S with a new type of thing, 'extensions', and let U be the class of everything whose existence the ontology acknowledged before this expansion. r is the relation of an 'intension' to its 'extension.' For each named 'intension' a we name its 'extension' a'. As remarked in Section 3, it seems obvious that we can do this selectively, to mix intensional and extensional semantics in one ontology - but this needs to be made precise.

6 Object Metamodeling

We can extend the semantics of this paper beyond the simple setting of set metamodeling. Let an *object type* of length k be a tuple $\tau \in \mathbb{N}^k$. We write k as $|\tau|$. Let L be a logic. Now define the logic L^{OM} to have axioms the axioms of L, plus the additional axioms $(a \approx_\tau O)$ and $(a \not\approx_\tau O)$, where O is a tuple of $|\tau|$ predicate names O_i, such that $O_i \in N_{P,\tau[i]}$. Call such a tuple an *object label* of type τ.

For any set D, let D^τ denote the set of $|\tau|$-tuples R such that the i-th element R_i of R is a subset of $D^{\tau[i]}$. Let an *object metamodeling interpretation* be a first-order structure $\mathcal{I}$ together with a relation $\approx_{\mathcal{I},\tau}$ for each τ, which is a local bijection between $\Delta^{\mathcal{I}}$ and $(\Delta^{\mathcal{I}})^\tau$. We may write it simply as $\mathcal{I}$. For an object label O of type τ, we let $O^{\mathcal{I}}$ denote the tuple in $(\Delta^{\mathcal{I}})^\tau$ such that $(O^{\mathcal{I}})_i = (O_i)^{\mathcal{I}}$. Define satisfaction in L^{OM} relative to an object metamodeling interpretation $\mathcal{I}$ by

1. $\mathcal{I} \models A$ for A an L axiom iff $\mathcal{I} \models_L A$
2. $\mathcal{I} \models (a \approx_\tau O)$ iff $a^{\mathcal{I}} \approx_{\mathcal{I},\tau} O^{\mathcal{I}}$
3. $\mathcal{I} \models (a \not\approx_\tau O)$ iff $a^{\mathcal{I}} \not\approx_{\mathcal{I},\tau} O^{\mathcal{I}}$

This can be used to state properties of an entire *structure* of semantic objects, such as a graph: the graph with vertex set $V^{\mathcal{I}}$ and edge set $E^{\mathcal{I}}$ (where $V \in N_{P,1}, E \in N_{P,2}$) can be treated as an object $g^{\mathcal{I}}$ using the axiom $(g \approx_\tau \langle V, E \rangle)$, where $\tau = \langle 1, 2 \rangle$. It can also be used to include in ontologies properties of roles, such as transitivity or connectness, by metamodeling the role (an object of type $\langle 2 \rangle$) and asserting that it belongs to $\mathsf{ConnectedRole}$, etc. Of course, the semantics by itself does not guarantee that $\mathsf{ConnectedRole}^{\mathcal{I}}$ is a class of connected roles, but we have a way of at least defining these meta-properties as ordinary predicates, not annotations or something else.

We prove the generalization of Theorem 1.

Say that a logic L has n-ary predicate equality (resp. inequality) if for every two predicates $P, Q \in N_{P,n}$ it has an axiom X that holds in a structure $\mathcal{I}$ exactly if $P^{\mathcal{I}} = Q^{\mathcal{I}}$ (resp. $P^{\mathcal{I}} \neq Q^{\mathcal{I}}$), which can be generated from (the binary encoding of) P, Q in polynomial time. We will denote these axioms $P \equiv Q$ (resp. $P \not\equiv Q$). For object labels O, O' of type τ, let $O \equiv O'$ be shorthand for the set of axioms $O_i \equiv_\tau O'_i$ for all $1 \leq i \leq |\tau|$. Say that L has *disjunctions of predicate inequalities* if, for every two τ-object labels O, O', there is an axiom $O \not\equiv O'$ which holds in $\mathcal{I}$ iff for at least one $1 \leq i \leq |\tau|$, $(O_i)^{\mathcal{I}} \neq (O'_i)^{\mathcal{I}}$.

Theorem 11. *Let L be a decidable logic and d a constant. Then L^{OM}-satisfiability - on sets S such that L has individual and n-ary predicate equality and inequality and disjunctions of inequalities, for every arity n of predicate appearing in S, and no type τ appears in S with $|\tau| > d$ - is decidable by an algorithm that runs in nondeterministic polynomial time and makes one call to an oracle for L-satisfiability.*

Proof. Like the proof of Theorem 1. Say that a type τ appears in a set of axioms S if some axiom $(a \approx_\tau O)$ appears. Say that a τ-object label O appears potentially in S if each component O_i appears in S. We extend the notion of identification X for S as follows:

1. for every individual name a and type τ appearing in S, and every object label O appearing potentially in S, exactly one of $(a \approx_\tau O)$ or $(a \not\approx_\tau O)$ is in X

2. for every pair a, b of individual names appearing in S, exactly one of $a = b$ or $a \neq b$ is in X

3. for every pair P, Q of n-ary predicate names appearing in S, exactly one of $P = Q$ or $P \neq Q$ is in X, for all n that are the arity of some predicate name appearing in S

4. if $(a \approx_\tau O), (b \approx_\tau O'), (a = b) \in S \cup X$, then $A \equiv_\tau B \subseteq S \cup X$

5. if $(a \approx_\tau O), (b \approx_\tau O'), (O \equiv_\tau O') \subseteq S \cup X$, then $a = b \in S \cup X$.

6. if $(a \approx_\tau O), (a = b), (O \equiv_\tau O') \in S \cup X$, then $(b \approx_\tau O') \in S \cup X$.

Let $L(X)$ denote the set of L-axioms in X.

We prove, as in Theorem 1, that S is L^{OM}-satisfiable iff there exists an identification X for S such that $L(S) \cup L(X)$ is L-satisfiable. Identifications are still of polynomial size in $|S|$, since only polynomially many types τ appear in S, and (thanks to the constant d) only polynomially many τ-object labels potentially appear. □

7 Related Work

Metamodeling has been tackled in the description logic literature from many angles. [14] is motivated by OWL Full, an ontology language with no built-in type distinction between individuals and classes. [5] explores higher-order description logics, which relate to normal DL's as higher-order logic does to first-order logic. Both of these works and their extensions differ from ours in style by using logics without explicit type separation. [10][11] discuss higher-order DL's with type separation, but with many types - classes of individuals, classes of classes of individuals, etc. [10] proves a relation between this kind of logic and the kind from [14]. [8] studies metamodeling in description logics in the DL-Lite family, which are mostly less expressive than the logics we consider here. [6] performs a passive kind of metamodeling, in which metamodeling information is extracted from an ontology that does not explicitly use any. [12] uses an individual 'as' a class without extending the base logic at all, associating a with $\exists t.\{a\}$ for some 'type' predicate t. This simulates a kind of intensional metamodeling.

Our approach is most similar to Motz et al. [16] and Severi [18], which uses axioms that assert the 'alignment' of an individual with a class, but distinguishes individual names from class names. However, the semantics in these papers is stricter than ours, and we discuss why this might not be wanted in section 4.

8 Conclusions

With this paper we hope to make the case that metamodeling can be simpler than it has often been thought. Our system does not assume acyclicity, and *does* assume extensionality - and we show that both of these features can be reversed (or mixed with their opposite) in a natural way by designing one's ontology to do so. Our opinion is that if these kind of choices can be made by the ontology designer rather than the logic designer, they should be. Results like Theorem 8, which show that an easy-to-work-with semantics agrees with a more constrained semantics for all appropriately-built ontologies, seem to us a promising line of research.

We have isolated what features of a logic L are needed to make our metamodeling results work, and they are minimal. Results such as Theorems 1 and 11 hold not only for DL's and rule-based languages, but for a broad range of formalisms including full higher-order logics, fixpoint logics, etc. - since these do have semantics in terms of first-order structures, they merely use them in a higher-order way.

Theorem 3 may seem somewhat silly, since the kind of algorithm it rules out is extremely restricted (a reduction using no specific axioms except equality and inequality statements). But given the huge variety of logics studied in KR, it is potentially insightful to find minimal sets of properties enabling some operation. We hope more powerful theorems of this type will come in the future.

Of course, before any kind of metamodeling can be incorporated into ontology-management software, we need optimized algorithms for specific logics. [15] has already done this for OWL and the semantics from [16]; adapting this to the non-wellfounded setting is probably easy.

We have mentioned the possibility of using both intensional and extensional metamodeling in one ontology, and using wellfoundedness assumptions in a local scope - we leave it for future work to devise an intended semantics for such situations, and to prove that an ontology designer using our framework can simulate it, aiming for results similar to Theorem 8.

References

[1] Suzi A. Aleksander, James P. Balhoff, Seth Carbon, J. Michael Cherry, et al. The gene ontology knowledgebase in 2023. *Genetics*, 224, 2023.

[2] Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather L. Butler, J. Michael Cherry, Allan Peter Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna E. Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, and Gerald M. Rubin. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29, 2000.

[3] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider. The description logic handbook. 2007.

[4] Eunsuk Chang and Javed Mostafa. The use of SNOMED CT, 2013-2020: a literature review. *Journal of the American Medical Informatics Association*, 28, 06 2021.

[5] Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Higher-order description logics for domain metamodeling. *Proceedings of the AAAI Conference on Artificial Intelligence*, 25:183–188, 08 2011.

[6] Birte Glimm, Sebastian Rudolph, and Johanna Völker. Integrated metamodeling and diagnosis in owl 2. In *International Workshop on the Semantic Web*, 2010.

[7] Stephan Grimm and Pascal Hitzler. A preferential tableaux calculus for circumscriptive ALCO. In *International Conference on Web Reasoning and Rule Systems*, 2009.

[8] Zhenzhen Gu and Songmao Zhang. DL-Lite Full: A sub-language of OWL 2 Full for powerful meta-modeling. In *Description Logics*, 2021.

[9] Pascal Hitzler, Markus Krötzsch, and Sebastian Rudolph. *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC Press, 2010.

[10] Martin Homola, Ján Kľuka, Vojtěch Svátek, and Miroslav Vacura. Towards typed higher-order description logics. In *Description Logics*, 2013.

[11] Martin Homola, Ján Kľuka, Vojtěch Svátek, and Miroslav Vacura. Typed higher-order variant of SROIQ - why not? In *Description Logics*, 2014.

[12] Adila Alfa Krisnadhi, Pascal Hitzler, and Krzysztof Janowicz. On the capabilities and limitations of OWL regarding typecasting and ontology design pattern views. In *OWL: Experiences and Directions*, 2015.

[13] Nuno Maia and Matteo Nizzardo. Identity and extensionality in Boffa set theory. *Philosophia Mathematica*, 2024.

[14] Boris Motik. On the properties of metamodeling in OWL. *J. Log. Comput.*, 17:617–637, 2005.

[15] Regina Motz. OWL extended with meta-modelling. In *ISW-LOD@IBERAMIA*, 2016.

[16] Regina Motz, Edelweis Rohrer, and Paula Severi. Reasoning for ALCQ extended with a flexible meta-modelling hierarchy. In *Joint International Conference of Semantic Technology*, 2014.

[17] Adam Rieger. An argument for Finsler-Aczel set theory. *Mind*, 109:241–253, 2000.

[18] Paula Severi and Edelweis Rohrer. Lightweight description logics with meta-modelling. https://www.cs.le.ac.uk/people/ps56/Submitted.pdf, 2016.