Conservativity of Boolean algebras with operators over semilattices with operators

A. Kurucz,1 Y. Tanaka,2 F. Wolter3 and M. Zakharyaschev4

1Department of Informatics
King’s College London, UK
agi.kurucz@kcl.ac.uk

2Faculty of Economics
Kyushu Sangyo University, Japan
ytanaka@ip.kyusyu-u.ac.jp

3Department of Computer Science
University of Liverpool, UK
wolter@liverpool.ac.uk

4Department of Computer Science
Birkbeck College London, UK

Department of Informatics
Faculty of Economics
University of Liverpool, UK
Birkbeck College London, UK

The problems considered in this paper originate in recent applications of large scale ontologies in medicine and other life sciences. The profile OWL 2 EL of the OWL 2 Web Ontology Language,1 used for this purpose, is based on the description logic $\mathcal{EL}$ [7]. The syntactic terms of $\mathcal{EL}$, called concepts, are interpreted as sets in first-order relational models. Concepts are constructed from atomic concepts and constants for the whole domain and empty set using intersection and existential restrictions of the form $\exists R.C$, $R$ a binary relation and $C$ a concept, which are understood as $\exists y (R(x, y) \land C(y))$. From a modal logic point of view, concepts are modal formulas constructed from propositional variables and the constants $\top$, $\bot$ using conjunction and diamonds. An $\mathcal{EL}$-theory is a set of inclusions (or implications) between such concepts, and the main reasoning problem in applications of $\mathcal{EL}$ in life sciences is to decide whether an $\mathcal{EL}$-theory entails a concept inclusion when interpreted over a class of relational structures satisfying certain constraints on its binary relations. Standard constraints in OWL 2 EL are transitivity and reflexivity, for which reasoning in $\mathcal{EL}$ is $\text{PTIME}$-complete, as well as symmetry and functionality, for which reasoning is $\text{EXPTIME}$-complete [1, 2].

As in modal logic, apart from reasoning over relational models, one can try to develop a purely syntactical reasoning machinery using a calculus. In other words, we can define a more general algebraic semantics for $\mathcal{EL}$: the underlying algebras are bounded meet-semilattices with monotone operators (SLOs, for short), constraints are given by equational theories of SLOs, and the reasoning problem is validity of quasi-equations in such equational theories. The resulting more general entailment problem is not necessarily complete with respect to the ‘intended’ relational semantics. This paper presents our initial results in an attempt to clarify which equational theories of SLOs are complete in this sense and which are not. We also prove that the completeness problem—given a finitely axiomatised equational theory of SLOs, decide whether it is complete with respect to the relational semantics—is algorithmically undecidable, which establishes a principle limitation regarding possible answers to our research question.

An $\mathcal{EL}$-equation is an expression of the form $\varphi \leq \psi$, where $\varphi$ and $\psi$ are terms that are built from variables $x_j$, $j \geq 1$, using meet $\wedge$, unary operators $f_i$, for $i \in I$, and constants 1 and 0. An $\mathcal{EL}$-theory, $T$, is a set of $\mathcal{EL}$-equations; and an $\mathcal{EL}$-quasi-equation is an expression of the form $(\varphi_1 \leq \psi_1) \land \cdots \land (\varphi_n \leq \psi_n) \rightarrow (\varphi \leq \psi)$, where the $\varphi_i \leq \psi_i$ and $\varphi \leq \psi$ are $\mathcal{EL}$-equations. The class of SLOs $\mathfrak{A} = (A, \wedge, 0, 1, f_i)_{i \in I}$ validating all equations in $T$ is the variety $\mathcal{V}(T)$. The ‘intended’ relational semantics of $\mathcal{EL}$ is given by $\mathcal{EL}$-structures $\mathfrak{A} = (\Delta, R_i)_{i \in I}$, which consist of a set $\Delta \neq \emptyset$ and binary relations $R_i$ on it. Every such $\mathfrak{A}$ gives rise to the complex algebra $\mathfrak{A}^+ = (2^S, f_i)_{i \in I}$ of $\mathfrak{A}$, where $2^\Delta$ is the full Boolean set algebra over $\Delta$ and $F_i(X) = \{ x \in \Delta \mid \exists y \in X \, x R_i y \}$, for $X \subseteq \Delta$. Complex algebras (CAs) are special cases of Boolean algebras with normal and $\lor$-additive

---

1http://www.w3.org/TR/owl2-overview/

1

\[ \text{http://www.w3.org/TR/owl2-overview/} \]
operators (BAOs, for short). The class of bounded distributive lattices with normal and $\lor$-additive operators is denoted by DLO. Slightly abusing notation (and remembering the signatures of DLOs and BAOs), we may assume that $CA \subseteq BAO \subseteq DLO \subseteq SLO$.

Given a class $C$ of SLOs, an $EL$-theory $T$ and a quasi-equation $q$, we say that $q$ follows from $T$ over $C$ and write $T \models_C q$ if $A \models q$, for every $A \in C$ with $A \models T$. An $EL$-theory $T$ is said to be $C$-conservative if $T \models_C q$ implies $T \models_{SLO} q$, for every quasi-equation $q$. We call $T$ complete if it is $CA$-conservative.

A standard way of establishing completeness of a modal logic is by showing that its axioms generate what Goldblatt [3] calls a ‘complex variety.’ This notion works equally well in the $EL$ setting: We say that an $EL$-theory $T$ is complex if every $A \in \mathbb{V}(T)$ is embeddable in some $\mathfrak{S}^+ \in CA$ validating $T$. The following theorem provides our main tool for investigating completeness of $EL$-theories:

**Theorem 1.** For every $EL$-theory $T$, $T$ is complex iff $T$ is complete iff $T$ is BAO-conservative.

The proof of this theorem uses the fact that all $EL$-equations correspond to Sahlqvist formulas in modal logic. Therefore, every $A \in BAO$ validating an $EL$-theory $T$ is embeddable into some $\mathfrak{S}^+ \in CA$ validating $T$. It also follows from the ‘correspondence’ part of Sahlqvist’s theorem that the class of $EL$-structures validating any $EL$-theory is first-order definable. For example,

- $x \leq f(x)$ defines reflexivity;
- $f(f(x)) \leq f(x)$ defines transitivity;
- $x \land f(y) \leq f(x \land y)$ defines symmetry;
- $f(x \land f(y) \leq f(x \land y)$ defines functionality;
- $f(x \land y) \land f(x \land z) \leq f(x \land y \land f(z))$ defines linearity over quasi-orders.

(We refer the reader to [6] for first steps towards a correspondence theory for $EL$.) In contrast to modal logic, however, the ‘completeness’ part of Sahlqvist’s theorem does not hold. The possibly simplest example of an incomplete $EL$-theory is $\{ f(x) \leq x \}$ (to see that this theory is not complex, it is enough to consider the SLO with three elements 0 < $a$ < 1 and the operation $f$ such that $f(0) = f(1) = 1$).

SLOs validating the reflexivity and transitivity equations above (but without 0 and 1 in the signature) have been studied by Jackson [4] under the name ‘closure semilattices’ (CSLs). He proves that every CSL is embeddable into a BAO validating reflexivity and transitivity. With a slight modification of his technique, we can obtain:

**Theorem 2.** The $EL$-theory $\{ x \leq f(x), f(f(x)) \leq f(x) \}$ is complete.

A more general completeness result has been proved by Sofronie-Stokkermans [9]:

**Theorem 3 ([9]).** Every $EL$-theory consisting of equations of the form $f_1 \ldots f_n(x) \leq f(x)$, $n \geq 0$, is complete.

This result implies that reflexivity or transitivity alone is also complete. Using modifications of Sofronie-Stokkermans’ techniques, we can also cover symmetry, functionality, and some combinations thereof.

**Theorem 4.** The following $EL$-theories are complete:

- $\{ x \land f(y) \leq f(x \land y) \}$ (symmetry);
- $\{ f(x) \land f(y) \leq f(x \land y) \}$ (functionality);
- $\{ x \leq f(x), f(f(x)) \leq f(x), x \land f(y) \leq f(f(x) \land y) \}$ (reflexivity, transitivity and symmetry).
In general, completeness is not preserved under unions of \( \mathcal{EL} \)-theories. For example:

**Theorem 5.** Neither the union \( \mathcal{T}_1 \) of reflexivity and functionality, nor the union \( \mathcal{T}_2 \) of symmetry and functionality is complete.

Interestingly, in both cases one can easily restore completeness by adding the equation \( f(x) \leq x \) to \( \mathcal{T}_1 \), and by adding \( f(f(x)) \leq x \) to \( \mathcal{T}_2 \). (Observe that these equations are consequences of \( \mathcal{T}_1 \) and \( \mathcal{T}_2 \) in modal logic.)

We also have a full picture of extensions of

\[
\mathcal{T}_{55} = \{ x \leq f(x), f(f(x)) \leq f(x), x \land f(y) \leq f(f(x) \land y) \},
\]

using that these equations axiomatise the well-known modal logic \( S5 \), and normal CSLs in [4]:

**Theorem 6.** The \( \mathcal{EL} \)-theory \( \mathcal{T}_{55} \cup \{ f(x) \land f(y) \leq f(x \land y) \} \) is incomplete. All other (countably infinitely many) extensions of \( \mathcal{T}_{55} \) are complete.

As a first step towards general completeness results, we note the following analogue of completeness preservation under fusions of modal logics [5]. We call \( \mathcal{T}_1 \cup \mathcal{T}_2 \) a fusion of \( \mathcal{EL} \)-theories \( \mathcal{T}_1 \) and \( \mathcal{T}_2 \) if the sets of the \( f_i \)-operators occurring in \( \mathcal{T}_1 \) and \( \mathcal{T}_2 \) are disjoint.

**Theorem 7.** The fusion of complete \( \mathcal{EL} \)-theories is also complete.

The proofs of Theorems 3 and 4 go via two steps: (1) by embedding any SLO validating \( \mathcal{T} \) into a DLO validating \( \mathcal{T} \), and then (2) by embedding this DLO into a BAO validating \( \mathcal{T} \), using various extensions of Priestley’s [8] DL-to-BA embedding to the operators \( f_i \). As concerns step (1), we have the following result:

**Theorem 8.** Every \( \mathcal{EL} \)-theory containing only equations where each variable occurs at most once in the left-hand side is DLO-conservative.

An interesting example, showing that the condition on the number of occurrences of variables in the left-hand side of equations in Theorem 8 cannot be dropped, is given by the \( \mathcal{EL} \)-theory

\[
\mathcal{T}_{S4.3} = \{ x \leq f(x), f(f(x)) \leq f(x), f(x \land y) \land f(x \land z) \leq f(x \land f(y) \land f(z)) \}.
\]

Observe first that \( \mathcal{T}_{S4.3} \) defines a relation which is reflexive, transitive and right-linear, that is, \( \forall x, y, z (R(x, y) \land R(x, z) \rightarrow R(y, z) \lor R(z, y)) \). The modal logic determined by this frame condition is known as \( S4.3 \), and the \( \mathcal{EL} \)-equations above axiomatise, if added to the equations for BAOs, the corresponding variety. However, one can show the following:

**Theorem 9.** \( \mathcal{T}_{S4.3} \) is not DLO-conservative.

*Proof.* Consider the quasi-equation

\[
\mathbf{q} = (f(x) \land y = x \land f(y)) \rightarrow (f(x) \land f(y) = f(x \land y))
\]

and the SLO \( \mathfrak{A} = \langle A, \land, 0, 1, f \rangle \), where

\[
A = \{ 0, a, b, c, d, e, 1 \},
\]

\[
a \land b = a \land c = b \land c = 0,
\]

\[
d = a \lor b, \ e = b \lor c, \ 1 = d \lor e,
\]

\[
f(a) = d, \ f(c) = e, \ \text{and} \ f(x) = x \ \text{for the remaining} \ x \in A.
\]

One can check that \( \mathcal{T}_{S4.3} \models_{DLO} \mathbf{q} \); on the other hand, \( \mathfrak{A} \models \mathcal{T}_{S4.3} \), \( \mathfrak{A} \not\models \mathbf{q} \), and so \( \mathcal{T}_{S4.3} \not\models_{SLO} \mathbf{q} \). \( \square \)

Finally, we analyse the completeness problem for \( \mathcal{EL} \)-theories from the algorithmic point of view and show that it is impossible to give an effective syntactic criterion for completeness:
Theorem 10. It is undecidable whether a finite set $T$ of $\mathcal{EL}$-equations is complete.

The proof of this result proceeds in two steps. First, we show the following by reduction of the undecidable halting problem for Turing machines:

Theorem 11. Triviality of finite sets of $\mathcal{EL}$-equations is undecidable; more precisely, no algorithm can decide, given a finite set $T$ of $\mathcal{EL}$-equations, whether $T \models_{\text{SLO}} 0 = 1$.

In the second step, we prove that, for every $\mathcal{EL}$-theory $T$, the following two conditions are equivalent:

- the fusion of $T$ and $\{ f(x) \leq x \}$ is complete;
- $T \models_{\text{SLO}} 0 = 1$.

Theorem 10 is then an immediate consequence of Theorem 11 and this equivalence.

References