Conservativity of Boolean algebras with operators over semilattices with operators

A. Kurucz,[1] Y. Tanaka,[2] F. Wolter[3] and M. Zakharyaschev[4]

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

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

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

! ! [4] Department of Computer Science
Birkbeck College London, UK
michael@dcs.bbk.ac.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 $\exists y\,(R(x,y) \wedge C(y))$. 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 PTime-complete, as well as symmetry and functionality, for which reasoning is 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 φ and ψ 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*, $\mathcal{T}$, is a set of $\mathcal{EL}$-equations; and an $\mathcal{EL}$-*quasi-equation* is an expression of the form $(\varphi_1 \leq \psi_1)\ \&\ \cdots\ \&\ (\varphi_n \leq \psi_n) \to (\varphi \leq \psi)$, 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 $\mathcal{T}$ is the variety $\mathsf{V}(\mathcal{T})$. The 'intended' relational semantics of $\mathcal{EL}$ is given by $\mathcal{EL}$-*structures* $\mathfrak{F} = (\Delta, R_i)_{i\in I}$, which consist of a set $\Delta \neq \emptyset$ and binary relations R_i on it. Every such $\mathfrak{F}$ gives rise to the *complex algebra* $\mathfrak{F}^+ = (2^\Delta, F_i)_{i\in I}$ of $\mathfrak{F}$, where 2^Δ is the full Boolean set algebra over Δ and $F_i(X) = \{x \in \Delta \mid \exists y \in X\, xR_iy\}$, for $X \subseteq \Delta$. Complex algebras (CAs) are special cases of *Boolean algebras with normal and* $\vee$-*additive*

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

operators (BAOs, for short). The class of *bounded distributive lattices with normal and* $\vee$*-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 $\mathcal{C}$ of SLOs, an $\mathcal{EL}$-theory $\mathcal{T}$ and a quasi-equation $\boldsymbol{q}$, we say that $\boldsymbol{q}$ *follows from* $\mathcal{T}$ *over* $\mathcal{C}$ and write $\mathcal{T} \models_{\mathcal{C}} \boldsymbol{q}$ if $\mathfrak{A} \models \boldsymbol{q}$, for every $\mathfrak{A} \in \mathcal{C}$ with $\mathfrak{A} \models \mathcal{T}$. An $\mathcal{EL}$-theory $\mathcal{T}$ is said to be $\mathcal{C}$*-conservative* if $\mathcal{T} \models_{\mathcal{C}} \boldsymbol{q}$ implies $\mathcal{T} \models_{\mathrm{SLO}} \boldsymbol{q}$, for every quasi-equation $\boldsymbol{q}$. We call $\mathcal{T}$ *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 $\mathcal{EL}$ setting: We say that an $\mathcal{EL}$-theory $\mathcal{T}$ is *complex* if every $\mathfrak{A} \in \mathsf{V}(\mathcal{T})$ is embeddable in some $\mathfrak{F}^+ \in \mathsf{V}(\mathcal{T})$. The following theorem provides our main tool for investigating completeness of $\mathcal{EL}$-theories:

Theorem 1. *For every* $\mathcal{EL}$*-theory* $\mathcal{T}$*,*

$$\mathcal{T} \text{ is complex} \quad \text{iff} \quad \mathcal{T} \text{ is complete} \quad \text{iff} \quad \mathcal{T} \text{ is BAO-conservative.}$$

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

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

(We refer the reader to [6] for first steps towards a correspondence theory for $\mathcal{EL}$.) In contrast to modal logic, however, the 'completeness' part of Sahlqvist's theorem does not hold. The possibly simplest example of an *incomplete* $\mathcal{EL}$-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 $f(a) = f(0) = 0$ and $f(1) = 1$).

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* $\mathcal{EL}$*-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* $\mathcal{EL}$*-theory consisting of equations of the form* $f_1 \dots f_n(x) \leq f(x)$*,* $n \geq 0$*, 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* $\mathcal{EL}$*-theories are complete:*

- $\{x \wedge f(y) \leq f(f(x) \wedge y)\}$ *(symmetry);*
- $\{f(x) \wedge f(y) \leq f(x \wedge y)\}$ *(functionality);*
- $\{x \leq f(x),\ f(f(x)) \leq f(x),\ x \wedge f(y) \leq f(f(x) \wedge y)\}$ *(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

$$\mathcal{T}_{\mathrm{S5}} = \{x \leq f(x),\, f(f(x)) \leq f(x),\, x \wedge f(y) \leq f(f(x) \wedge y)\},$$

using that these equations axiomatise the well-known modal logic $\mathcal{S}5$, and *normal* CSLs in [4]:

Theorem 6. *The $\mathcal{EL}$-theory $\mathcal{T}_{\mathrm{S5}} \cup \{f(x) \wedge f(y) \leq f(x \wedge y)\}$ is incomplete. All other* (*countably infinitely many*) *extensions of $\mathcal{T}_{\mathrm{S5}}$ 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

$$\mathcal{T}_{\mathrm{S4.3}} = \{x \leq f(x),\, f(f(x)) \leq f(x),\, f(x \wedge y) \wedge f(x \wedge z) \leq f(x \wedge f(y) \wedge f(z))\}.$$

Observe first that $\mathcal{T}_{\mathrm{S4.3}}$ defines a relation which is reflexive, transitive and right-linear, that is, $\forall x, y, z(R(x,y) \wedge R(x,z) \to R(y,z) \vee R(z,y))$. The modal logic determined by this frame condition is known as $\mathcal{S}4.3$, 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. *$\mathcal{T}_{\mathrm{S4.3}}$ is not DLO-conservative.*

Proof. Consider the quasi-equation

$$\boldsymbol{q} = \big(f(x) \wedge y = x \wedge f(y)\big) \;\to\; \big(f(x) \wedge f(y) = f(x \wedge y)\big)$$

and the SLO $\mathfrak{A} = (A, \wedge, 0, 1, f)$, where

$$\begin{aligned}
&A = \{0, a, b, c, d, e, 1\},\\
&a \wedge b = a \wedge c = b \wedge c = 0,\\
&d = a \vee b,\ e = b \vee c,\ 1 = d \vee e,\\
&f(a) = d,\ f(c) = e,\ \text{and } f(x) = x \text{ for the remaining } x \in A.
\end{aligned}$$

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

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 $\mathcal{T}$ 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 $\mathcal{T}$ of $\mathcal{EL}$-equations, whether $\mathcal{T} \models_{\mathrm{SLO}} 0 = 1$.*

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

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

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

References

[1] F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. of IJCAI*, 2005.

[2] F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope further. In *Proc. of OWLED Workshop on OWL: Experiences and Directions*, 2008.

[3] R. Goldblatt. Varieties of complex algebras. *Annals of Pure and Applied Logic*, 38:173–241, 1989.

[4] M. Jackson. Semilattices with closure. *Algebra Universalis*, 52:1–37, 2004.

[5] M. Kracht and F. Wolter. Properties of independently axiomatizable bimodal logics. *Journal of Symbolic Logic*, 56(4):1469–1485, 1991.

[6] A. Kurucz, F. Wolter, and M. Zakharyaschev. Islands of tractability for relational constraints: towards dichotomy results for the description logic $\mathcal{EL}$. In L. Beklemishev, V. Goranko, and V. Shehtman, editors, *Advances in Modal Logic, Volume 8*, pages 271–291. College Publications, 2010.

[7] C. Lutz and F. Wolter. Mathematical logic for life science ontologies. In *WoLLIC*, pages 37–47, 2009.

[8] H. Priestley. Representations of distributive lattices by means of ordered Stone spaces. *Bull. London Math. Soc.*, 2:186–190, 1970.

[9] V. Sofronie-Stokkermans. Locality and subsumption testing in $\mathcal{EL}$ and some of its extensions. In C. Areces and R. Goldblatt, editors, *Advances in Modal Logic, Volume 7*, pages 315–339. College Publications, 2008.