Embedding Approaches to Combining Rules and Ontologies into Autoepistemic Logic

Jos de Bruijn
Faculty of Computer Science
Free University of Bozen-Bolzano
Piazza Domenicani 3, Bolzano, Italy
debruijn@inf.unibz.it

Thomas Eiter and **Hans Tompits**
Institute of Information Systems
Vienna University of Technology
Favoritenstraße 9-11, Vienna, Austria
{eiter|tompits}@kr.tuwien.ac.at

Abstract

The combination of rules and ontologies has a central role in the ongoing development of the Semantic Web. In previous work, *autoepistemic logic* (AEL) was advocated as a uniform host formalism to study different such combinations, enabling comparisons on a common basis. In this paper, we continue this line of research and investigate different embeddings of major proposals to combine rules and ontologies into first-order autoepistemic logic (FO-AEL). In particular, we present embeddings for dl-programs, r-hybrid knowledge bases, and hybrid MKNF knowledge bases, which are representatives of different combination types. We study the embeddings in the context of FO-AEL under the standard-names assumption, but we also discuss variants using the any- and all-names semantics. Our results provide interesting insights into the properties of the discussed combination formalisms.

Introduction

The Semantic Web vision, in which rules and ontologies play a prominent role as tools for describing content at the semantic level, has raised interest in KR formalisms that incorporate both rules and ontologies. In particular, combining logic programs and description logic ontologies (e.g., OWL DL) has received significant attention.

As logic programming and description logics are not based on common grounds and have quite different features (e.g., open vs. closed world and domain), overcoming the impedance mismatch between them is a nontrivial issue that has led to many proposals for a semantics of a "hybrid" knowledge base (KB) that consists of a rules and an ontology part; see de Bruijn *et al.* (2006), Rosati (2005), Motik & Rosati (2007b) for surveys and discussions.

This raises the desire for a uniform formalism that, in addition to enabling a seamless combination of rules and ontologies in a genuine way, can also serve as a *lingua franca* into which various existing combination proposals can be embedded. Using such a formalism, one can compare different approaches on a common basis, and derive properties of them via the properties of the host formalism.

In previous work (de Bruijn *et al.* 2007), autoepistemic logic (AEL) was advocated as a uniform formalism for combining rules and ontologies, and various embeddings of non-monotonic logic programs into first-order AEL (FO-AEL) have been studied; properties of these embeddings with respect to a naive combination of a set of rules under stable semantics and an ontology formulated in first-order logic have been analyzed. In the present paper, we go further and investigate embeddings of major proposals to combining rules and ontologies into FO-AEL. As many other works, we focus on ontologies that are effectively subsets of first-order logic, like many *classical* core description logics.

In particular, we consider non-monotonic dl-programs (Eiter *et al.* 2004), r-hybrid knowledge bases (also known as $\mathcal{DL}+log$) (Rosati 2005; 2006), and hybrid MKNF knowledge bases (Motik & Rosati 2007a; 2007b); these are prominent representatives of the three combination types (de Bruijn *et al.* 2006) that foster a *loose coupling* (via an inference interface), a *hybrid integration* (where the predicates of the rules and the ontology parts are strictly separated), and a *full integration* (where no distinction of predicates is made), respectively. Note that SWRL (Horrocks *et al.* 2005), a rule-based extension of OWL DL, essentially corresponds to r-hybrid knowledge bases but without negation in the rules (Rosati 2005).

Besides AEL, other unifying formalisms have been proposed to study combinations of rules and ontologies, notably MKNF (Motik & Rosati 2007a) and quantified equilibrium logic (de Bruijn *et al.* 2007b), which is suited for reasoning about knowledge-base equivalences by its link to intermediate logics. Compared to these formalisms, AEL has some attractive features that make its consideration worthwhile. First, AEL (in particular, its propositional fragment) is a premier formalism of non-monotonic reasoning that has been widely studied, en par with Reiter's default logic and McCarthy's circumscription. Second, the language is "plain" as it has a single modal operator, and also the definition of the semantics is quite intuitive—e.g., AEL can be characterized in terms of classical (i.e., non-modal) consequence using Moore's fixpoint equation (Moore 1985) or in terms of maximal $\mathbf{S5}$-models (Levesque 1990). Finally, many variants of AEL have been considered, and its relationship to other non-monotonic modal logics is well-studied (Marek & Truszczyński 1993; Rosati 1997). This provides a rich stock of knowledge to draw from in the analysis of combinations of rules and ontologies.

In this paper, we consider a variant of FO-AEL under the standard-names assumption (SNA), extending the treatment of Levesque (1990) with function symbols. Under this assumption, the any-name and all-name semantics of Konolige (1991) coincide and, as we show, the characterizations of stable expansions by Konolige (1991), Moore (1985), and Levesque (1990) are equivalent in this setting.

Our main results are summarized as follows.

• We provide a number of results on the embeddability of dl-programs and r-hybrid and hybrid MKNF knowledge bases into FO-AEL. All these formalisms can be faithfully embedded into FO-AEL such that models of a hybrid KB correspond with stable expansions. However, in some cases the embeddings are not necessarily modular in the sense of Janhunen (1999), i.e., an embedding on a formula-by-formula basis is infeasible. Specifically, we establish the following.

– We give an embedding of dl-programs under the weak answer set semantics into FO-AEL that is polynomially computable, faithful, and modular (PFM). For the strong answer-set semantics, a non-modular embedding is obtained by combining results on embedding MKNF into FO-AEL and results by Motik and Rosati on embedding dl-programs into MKNF.

– We present a simple PFM-embedding of r-hybrid KBs into FO-AEL. Roughly speaking, it combines the ontology with a straightforward embedding of a logic program, using additional axioms that ensure classical interpretation of the ontology predicates.

– For hybrid MKNF KBs, we consider embeddings for different fragments of the formalism. We show that for the general case no FM-embedding exists, and present a faithful yet non-modular (and non-polynomial) embedding into FO-AEL, generalizing earlier work by Rosati (1997). We base it on a PFM-embedding for MKNF clause sets, generalizing earlier results by Lifschitz & Schwarz (1993).

• Finally, we briefly discuss differences between the standard-names semantics and the any- respectively all-names semantics (Konolige 1991), extend Moore's fixpoint characterization of stable expansions in terms of classical entailment for the first-order case with standard names, and discuss variants of the embeddings under the any- and all-name semantics.

These results provide us with new interesting insights into the semantic properties of the three combination approaches. Given that logic programming possesses several PFM-embeddings into FO-AEL (de Bruijn *et al.* 2007), the non-existence of FM-embeddings for hybrid MKNF KBs into FO-AEL—and in fact the concrete examples showing this failure—provide some further evidence that the interaction of rules and axioms in these formalisms is handled in a way that goes beyond separate consideration of the two parts and gluing them together. On the other hand, the results on r-hybrid KBs may, oversimplifying, be phrased as "r-hybrid KB = LP + ontology + CI axioms", revealing that $\mathcal{DL}+log$ is an elegant and effective extension of the naive basic combination "hybrid KB = LP + ontology" (de Bruijn *et al.* 2007), where "+" means union of sets of straightforwardly embedded parts.

The paper is structured as follows. We first review first-order autoepistemic logic, logic programs, and earlier results on embedding logic programs into FO-AEL. For each of the formalisms of dl-programs, r-hybrid KBs, and hybrid MKNF KBs, we then proceed with reviews of the formalisms, with statements of our results, as described in the above, and with discussion of implications of the results. After that we discuss the any- and all-name semantics, and implications on the embeddings when dropping the standard names assumption. In the final section we provide discussion and some conclusions.

Preliminaries

Autoepistemic logic (AEL) (Moore 1985) is the logic of belief of an introspective agent, i.e., an agent with knowledge about its own beliefs. It extends classical logic with a non-monotonic modal belief operator: $\mathsf{L}\phi$ intuitively means that ϕ is believed (or known) by the agent, and amounts to the non-monotonic version of the modal logic $\mathbf{k45}$ (Marek & Truszczyński 1993). In the absence of the belief operator, AEL coincides with classical logic. Consistency and consequence in AEL are defined based on the notion of *stable expansion*, which is informally an acceptable belief state of the agent that is defined in terms of a fixpoint equation; the non-monotonicity of AEL stems from this equation.

We consider in this paper first-order autoepistemic logic (FO-AEL) (Konolige 1991) under the standard-names assumption (Levesque 1990). We review two characterizations of stable expansions, an entailment-based characterization involving a fixpoint equation (due to Konolige (1991)) and a model-based characterization in terms of $\mathbf{S5}$ structures subject to a maximality condition (due to Levesque (1990)).

First-order Logic We consider first-order logic (FOL) with equality. An FO language $\mathcal{L}$ is defined over a signature $\Sigma_{\mathcal{L}} = (\mathcal{F}, \mathcal{P})$, where $\mathcal{F}$ and $\mathcal{P}$ are disjoint countable sets of *function* and *predicate symbols*, respectively, each having a non-negative arity n. Function symbols with arity 0 are called *constants*. $\mathcal{V}$ is a countably infinite set of *variable symbols*. Terms and atomic formulas (atoms) are constructed as usual; $\top$ and $\bot$ are atoms. Ground terms are also called *names* and $\mathcal{N}_\Sigma$ is the set of names of a given signature Σ; we write $\mathcal{N}$ when the signature is clear from the context. Complex formulas are constructed as usual from atoms using the connectives $\neg$, $\wedge$, $\vee$, $\supset$, $\equiv$ and quantifiers $\exists$, $\forall$. *Literals* are atoms or negated atoms. The universal closure of a formula ϕ is denoted by $(\forall)\ \phi$. $\mathcal{L}_{gl}$ (resp., $\mathcal{L}_{ga}$) is the restriction of $\mathcal{L}$ to ground literals (resp., atoms). An *FO theory* $\Phi \subseteq \mathcal{L}$ is a set of formulas with no free variables.

As usual, an *interpretation* of a language $\mathcal{L}$ is defined as a tuple $\mathcal{I} = \langle U, \cdot^I \rangle$, where U is the domain and $\cdot^I$ is a mapping taking care of the assignment of function and predicate symbols in U; $\cdot^I$ extends to ground terms in the usual way.

A *variable assignment* B for $\mathcal{I}$ is a mapping that assigns an element $B(x) \in U$ to every variable $x \in \mathcal{V}$. Satisfaction of an objective formula ϕ in $\mathcal{I}$ relative to B, denoted $\mathcal{I}, B \models \phi$, is as usual; and $\mathcal{I}$ is a *model* of ϕ, denoted $\mathcal{I} \models \phi$, if

$\mathcal{I} \models_\Gamma \phi$ for every variable assignment B of $\mathcal{I}$.

A variable assignment B' is an *x-variant* of B if $B(y) = B'(y)$ for every variable $y \in \mathcal{V}$ such that $y \neq x$.

The *standard-names assumption* (SNA) is defined as follows: a signature Σ is an *SNA signature* if it includes a countably infinite set of constant symbols and the two-place predicate symbol '$\approx$', which is written in infix notation. An interpretation $\mathcal{I} = \langle U, \cdot^I \rangle$ is an *SNA interpretation* if (a) $\mathcal{I}$ is an interpretation of some language $\mathcal{L}$ with an SNA signature Σ, (b) $U = \mathcal{N}_\Sigma$, (c) $t^I = t$ for any name $t \in \mathcal{N}_\Sigma$, and (d) $\approx^I$ is a congruence relation over U. In the sequel, we consider only SNA interpretations, unless mentioned otherwise.

Equality under the SNA behaves quite differently from the usual first-order equality, hence the symbol '$\approx$'. If ψ is a formula (resp., theory) not containing '$\approx$', then $\psi^\approx$ results from ψ by replacing every occurrence of '$=$' with '$\approx$'.

First-order Autoepistemic Logic The FO-AEL language $\mathcal{L}_\mathsf{L}$ relative to a language $\mathcal{L}$ allows for the belief operator L in front of arbitrary subformulas. For example, $\exists x.\mathsf{L}p(x)$ is an FO-AEL formula: the use of free variables in modal atoms (also known as *quantifying-in*) is allowed. An *FO-AEL theory* $\Phi \subseteq \mathcal{L}_\mathsf{L}$ is a set of formulas with no free variables.

A formula of the form $\mathsf{L}\phi$ is called a *modal atom*. A formula without modal atoms is an *objective* formula. *Standard autoepistemic logic* is FO-AEL without variables.

Example 1. *Consider the theory* $\Phi = \{\neg\mathsf{L}win \supset lose, \neg\mathsf{L}lose \supset win, \mathsf{L}win \vee \mathsf{L}lose \supset bet\}$. *Intuitively, it may capture an agent's attitude that disbelief in a team winning a game implies that it loses, and disbelief in losing that it wins; furthermore, that betting is advisable if one believes in either winning or losing.*

Example 2. *Consider the theory* Φ *given by the set*

$$\{\forall x.drink(x) \supset beverage(x), \forall x.drink(x) \wedge \neg\mathsf{L}\neg alcoholic(x) \supset alcoholic(x), drink(mojito)\}.$$

The intuitive meaning of the first and third formula is clear; the second expresses that a drink is alcoholic, unless the opposite is known.

An *autoepistemic interpretation* is a pair $\langle \mathcal{I}, \Gamma \rangle$, where $\mathcal{I} = \langle U, \cdot^I \rangle$ is an interpretation and $\Gamma \subseteq \mathcal{L}_\mathsf{L}$ is a *belief set*. Satisfaction of objective formulas ϕ in $\langle \mathcal{I}, \Gamma \rangle$ relative to a variable assignment B of $\mathcal{I}$, denoted $\mathcal{I}, B \models_\Gamma \phi$, coincides with $\mathcal{I}, B \models \phi$; for modal atoms $\mathsf{L}\phi$, satisfaction is defined by

$$\mathcal{I}, B \models_\Gamma \mathsf{L}\phi \text{ iff } B(\phi) \in \Gamma, \tag{1}$$

where $B(\phi)$ is obtained from ϕ by replacing every variable x with $B(x)$ (recall that $B(x)$ is a ground term, by the SNA). Satisfaction $\mathcal{I}, B \models_\Gamma \phi$ of other formulas ϕ is then structurally defined as usual (Konolige 1991). $\langle \mathcal{I}, \Gamma \rangle$ is a *model* of ϕ, denoted $\mathcal{I} \models_\Gamma \phi$, if $\mathcal{I}, B \models_\Gamma \phi$ for every variable assignment B of $\mathcal{I}$. This extends to sets of formulas in the usual way. A theory $\Phi \subseteq \mathcal{L}_\mathsf{L}$ *entails* a formula $\phi \in \mathcal{L}_\mathsf{L}$ with respect to a belief set Γ, denoted $\Phi \models_\Gamma \phi$, if for every interpretation $\mathcal{I}$ such that $\mathcal{I} \models_\Gamma \Phi$, $\mathcal{I} \models_\Gamma \phi$.

A belief set $T \subseteq \mathcal{L}_\mathsf{L}$ is a *stable expansion* of a theory $\Phi \subseteq \mathcal{L}_\mathsf{L}$ if $T = \{\phi \mid \Phi \models_T \phi\}$. T_{ogl} and T_{oga} denote the restrictions of T to objective ground literals and atoms, respectively, i.e., $T_{ogl} = T \cap \mathcal{L}_{gl}$ and $T_{oga} = T \cap \mathcal{L}_{ga}$. A formula ϕ is an *autoepistemic consequence* of Φ if ϕ is included in every stable expansion of Φ.

Example 3. *Reconsider the theory* Φ *from Example 1. It has the following two stable expansions:*

$T_1 = \{lose, bet, \mathsf{L}lose, \mathsf{L}bet, \neg\mathsf{L}win, \mathsf{L}win \vee \mathsf{L}lose, \mathsf{L}\mathsf{L}lose, \ldots\}$;
$T_2 = \{win, bet, \mathsf{L}win, \mathsf{L}bet, \neg\mathsf{L}lose, \mathsf{L}win \vee \mathsf{L}lose, \mathsf{L}\mathsf{L}win, \ldots\}$.

Φ *has an autoepistemic consequence,* bet. *In fact, this is the only objective atom that is a consequence of* Φ.

Example 4. *The theory* Φ *from Example 2 has a single stable expansion* T*, given by*

$$\{drink(mojito), \mathsf{L}drink(mojito), alcoholic(mojito), \mathsf{L}alcoholic(mojito), \neg\mathsf{L}\neg alcoholic(mojito), \ldots\}.$$

For every standard name $t \neq mojito$*, it contains the formulas* $\neg\mathsf{L}(\neg)drink(t)$, $\neg\mathsf{L}(\neg)beverage(t)$, *and* $\neg\mathsf{L}(\neg)alcoholic(t)$, *hence also no belief about* t *being a drink, a beverage, or alcoholic.*

The following characterization of stable expansions is a straightforward generalization of de Bruijn *et al.* (2007, Proposition 1).

Proposition 1. *Given a base set* $\Phi \subseteq \mathcal{L}_\mathsf{L}$ *with only objective literals in the context of modal atoms, a set of objective ground literals* $\Gamma \subseteq \mathcal{L}$*, and a belief set* $T \subseteq \mathcal{L}_\mathsf{L}$ *such that* $T_{ogl} = \Gamma$,

$$T = \{\phi \in \mathcal{L} \mid \Phi \models_\Gamma \phi\} \text{ iff } T \text{ is a stable expansion of } \Phi.$$

For the analysis of MKNF, it is convenient to use an alternative semantic characterization of stable expansions, extending Levesque (1990), in terms of AEL models, which are **S5** structures subject to a maximality condition. In more detail, we consider pairs $\langle \mathcal{I}, M \rangle$, where $\mathcal{I}$ is an interpretation and M is a set of interpretations. Satisfaction of objective formulas ϕ relative to a variable assignment B of $\mathcal{I}$ coincides with $\mathcal{I} \models \phi$, and of modal atoms is defined as

$$\langle \mathcal{I}, M \rangle, B \models \mathsf{L}\phi \text{ iff } \langle \mathcal{I}', M \rangle, B \models \phi, \text{ for all } \mathcal{I}' \in M. \tag{2}$$

Satisfaction of other formulas and $\langle \mathcal{I}, M \rangle \models \phi$ are then defined as usual. A set M of interpretations satisfies a formula ϕ, denoted $M \models \phi$, if $\langle \mathcal{I}, M \rangle \models \phi$, for every $\mathcal{I} \in M$, and analogous for theories. M is an *AEL model* of a theory $\Phi \subseteq \mathcal{L}_\mathsf{L}$ if $M = \{\mathcal{I} \mid \langle \mathcal{I}, M \rangle \models \Phi\}$.

Proposition 2. *A belief set* $T \subseteq \mathcal{L}_\mathsf{L}$ *is a stable expansion of an FO-AEL theory* $\Phi \subseteq \mathcal{L}_\mathsf{L}$ *iff* Φ *has an AEL model* M *such that* $T = \{\phi \mid M \models \phi\}$.

This result follows from Proposition 12 and a straightforward extension of Levesque (1990, Theorem 3.16). Note that the empty set of interpretations corresponds to the inconsistent expansion.

Logic Programs We consider disjunctive logic programs with strong negation (written '$\neg$') under the answer-set semantics (Gelfond & Lifschitz 1991). A *logic program* P of a language $\mathcal{L}$ with at least one constant is a set of rules

$$h_1 \mid \ldots \mid h_l \leftarrow b_1, \ldots, b_m, \mathit{not}\ c_1, \ldots, \mathit{not}\ c_n \tag{3}$$

where $h_1, \ldots, h_l, b_1, \ldots, b_m, c_1, \ldots, c_n$ are literals, with m, $n \geq 0$ and $l \geq 1$, such that $h_1, \ldots, h_l$ are equality-free and no b_i or c_i contains contains both '¬' and '='. We define $H(r) = \{h_1, \ldots, h_l\}$, $B^+(r) = \{b_1, \ldots, b_m\}$, and $B^-(r) = \{c_1, \ldots, c_n\}$ as the sets of *head literals*, *positive body literals*, and *negated body literals* of r, respectively. If $B^-(r) = \emptyset$, then r is *positive*. If every variable in r occurs in $B^+(r)$, then r is *safe*. Likewise for programs P.

We denote by $gr(P)$ the grounding of P with names in $\mathcal{N}_{\Sigma_\mathcal{L}}$. A *Herbrand interpretation* I of $\mathcal{L}$ is a consistent set of ground literals. I is a *model* of a positive program P if $\top \in I$, $\bot \notin I$, $n = n \in I$ for every $n \in \mathcal{N}_{\Sigma_\mathcal{L}}$, and for every rule $r \in gr(P)$, $B^+(r) \subseteq I$ implies $H(r) \cap I \neq \emptyset$.

The *reduct* of a logic program P with respect to I, denoted P^I, is obtained from $gr(P)$ by deleting (i) each rule r with $B^-(r) \cap I \neq \emptyset$, and (ii) $not\ c$ from the body of every remaining rule r such that $c \in B^-(r)$. If I is a model of P^I and for every model $I' \neq I$ of P^I holds that $I' \not\subseteq I$, then I is a *stable model* of P.

Embeddings An embedding τ, as discussed in our paper, takes as input a logic program, a theory, or the combination of a program and a theory; the output is a theory of FO-AEL.

An embedding τ is *faithful* if, for each input $\imath$, there is a one-to-one correspondence between models of $\imath$ in the input formalism and stable expansions of $\tau(\imath)$.

An embedding τ is *modular* if, for any two inputs $\imath_1$ and $\imath_2$, $\tau(\imath_1 \cup \imath_2) = \tau(\imath_1) \cup \tau(\imath_2)$. Furthermore, τ is *signature-modular* if, for any two inputs $\imath_1$ and $\imath_2$ with the same signature Σ, $\tau(\imath_1 \cup \imath_2) = \tau(\imath_1) \cup \tau(\imath_2)$.

Embedding Logic Programs into FO-AEL We briefly recapitulate (and slightly extend) previous results on embedding non-ground logic programs into FO-AEL (de Bruijn *et al.* 2007). We consider two embeddings, called HP and EH. In the HP embedding, the modal operator is not applied to positive literals and disjunction-free positive rules are embedded as Horn formulas; HP stands for "*H*orn for *P*ositive rules". In the EH embedding, the modal operator is applied to both positive body and head literals; EH stands for "*E*pistemic rule *H*eads".

To cope with disjunction in a manner consistent with the answer-set semantics, the HP embedding requires the so-called *positive introspection* (PI) *axioms*. With PIA_Σ we denote the set of axioms

PIA $\quad \alpha \supset \mathsf{L}\alpha$, for every objective ground atom α of Σ.

Definition 1. *Let P be a logic program and r a rule of form* (3)*. Then,*

$\tau_{HP}(r) = (\forall)\ \bigwedge_i b_i \wedge \bigwedge_j \neg \mathsf{L} c_j \supset \bigvee_k h_k$,

$\tau_{EH}(r) = (\forall)\ \bigwedge_i (b_i \wedge \mathsf{L} b_i) \wedge \bigwedge_j \neg \mathsf{L} c_j \supset \bigvee_k (h_k \wedge \mathsf{L} h_k)$,

$\tau_{HP}(P) = \{\tau_{HP}(r) \mid r \in P\} \cup PIA_{\Sigma_P}$*, and*

$\tau_{EH}(P) = \{\tau_{EH}(r) \mid r \in P\}$.

Notice that the HP embedding is signature-modular and the EH embedding is modular. The next proposition follows straightforwardly from results by de Bruijn *et al.* (2007) and Lifschitz & Schwarz (1993), given the SNA. Note that the HP embedding is not faithful for logic programs with strong negation, as discussed by Lifschitz & Schwarz (1993).

Proposition 3. *A Herbrand interpretation I is a stable model of a logic program P iff there is a consistent stable expansion T of $\tau_{EH}(P)$ such that $I = T \cap \mathcal{L}_{gl}$. If P does not contain strong negation, then there is a consistent stable expansion T' of $\tau_{HP}(P)$ such that $I = T' \cap \mathcal{L}_{ga}$.*

The proposition establishes a one-to-one correspondence between stable models of logic programs and consistent stable expansions of their embeddings. It follows that there is a correspondence between notions of consistency (existence of a stable model versus existence of a consistent stable expansion) as well as various notions of entailment (e.g., cautious entailment versus autoepistemic consequence).

Embedding dl-programs

We consider dl-programs (Eiter *et al.* 2004) with function symbols of arbitrary arity and disjunction in rule heads. Moreover, the classical theories with which the logic programs are combined need not be from a particular description logic, but may be arbitrary first-order theories and queries may be arbitrary formulas, including conjunctive queries (Eiter *et al.* 2007). We have one restriction: the classical theory and the logic program must have the same (infinite) set of names in their respective signatures. Eiter *et al.* (2004) defined both a weak and a strong answer-set semantics for dl-programs. We consider here the weak answer-set semantics and discuss possible embeddings of the strong answer-set semantics later on.

Let $\mathcal{L}^\Phi$ and $\mathcal{L}^P$ be FO languages with respective SNA signatures $\langle \mathcal{F}, \mathcal{P}_\Phi \rangle$ and $\langle \mathcal{F}, \mathcal{P}_P \rangle$ such that $\mathcal{P}_\Phi \cap \mathcal{P}_P = \emptyset$. Symbols in $\mathcal{P}_\Phi$ (resp., $\mathcal{P}_P$) are called *classical* (resp., *rule*) *predicates*. A *dl-atom* has the form

$$DL[S_1 op_1 p_1, \ldots, S_m op_m p_m; Q](t_1, \ldots, t_n), \qquad (4)$$

where $S_i \in \mathcal{P}_\Phi$ and $p_i \in \mathcal{P}_P$ are k-ary predicate symbols, $op_i \in \{\uplus, \cup\!\!\!\!-, \cap\!\!\!\!-\}$, Q is an n-ary classical predicate or a formula in $\mathcal{L}^\Phi$ with n free variables, and $t_1, \ldots, t_n$ are terms. A *dl-rule* is like a rule of the form (3), except that the head literals $h_1, \ldots, h_l$ are equality-free literals of $\mathcal{L}^P$ and the body literals $b_1, \ldots, c_n$ are literals of $\mathcal{L}^P$ or dl-atoms. A *dl-program* is a pair (Φ, P), where $\Phi \subseteq \mathcal{L}^\Phi$ is a finite FO theory and P is a set of dl-rules.

Example 5. *Consider the dl-program* (Φ, P)*, where we have* $\Phi = \emptyset$ *and*

$$P = \{a(t);\ b(t) | b'(t) \leftarrow DL[A \uplus a; A](t); \\ c(x) \leftarrow b(x), not\ DL[A \cup\!\!\!\!- a; A](x)\}.$$

P contains a fact $a(t)$, and the first rule informally says that either $b(t)$ or $b(t')$ is true if after adding the contents of a to A, $A(t)$ is true from Φ; the second rules says that $c(x)$ is true if $b(x)$ is true and, after adding the contents of a to the complement of A, $A(x)$ is not provable from Φ.

Formally, the semantics of dl-programs is as follows. We denote the grounding of P by $gr(P)$ and the set of dl-atoms in a rule r (resp., set of rules P) by $DL(r)$ (resp., $DL(P)$).

A Herbrand interpretation I of $\mathcal{L}^P$ is a *model* of a literal l under Φ if $l \in I$ and of a ground dl-atom of form (4) under Φ if $\Phi \cup \bigcup_{i=1}^m A_i(I) \models Q(t_1, \ldots, t_n)$, where

- $A_i(I) = \{S_i(\vec{e}) \mid p_i(\vec{e}) \in I\}$, for $op_i = \uplus$,

- $A_i(I) = \{\neg S_i(\vec{e}) \mid p_i(\vec{e}) \in I\}$, for $op_i = \uplus$, and
- $A_i(I) = \{\neg S_i(\vec{e}) \mid p_i(\vec{e}) \notin I\}$, for $op_i = \Cap$,

where $\vec{e} = e_1, \ldots, e_n$ are ground terms.

Now, I is a *model* of a ground dl-rule r under Φ if I is a model of some $h_i \in \{h_1, \ldots, h_l\}$ whenever I is a model of all $b_i \in \{b_1, \ldots, b_m\}$ and is not a model of any $c_i \in \{c_1, \ldots, c_n\}$ under Φ. I is a model of a dl-program (Φ, P) if I is a model of every $r \in gr(P)$ under Φ.

For a dl-program (Φ, P), the *weak dl-transform* of P relative to Φ and a model I of P, denoted wP^I_Φ, is the logic program obtained from $gr(P)$ by deleting

- each $r \in gr(P)$ such that either I is not a model of some $\alpha \in B^+(r) \cap DL(r)$, or a model of some $\alpha \in B^-(r)$, and
- all literals in $B^-(r) \cup (B^+(r) \cap DL(r))$ from each remaining $r \in gr(P)$.

If I is an answer set of the logic program wP^I_Φ, then I is a *weak answer set* of KB.

Example 6. *The program (Φ, P) from Example 6 has two weak answer sets: $\{a(t), b(t), c(t)\}$ and $\{a(t), b'(t)\}$.*

Embedding the Weak Answer-Set Semantics We now generalize the embedding τ_{EH} to the case of dl-programs under the weak answer-set semantics.

Definition 2. *Let (Φ, P) be a dl-program and let $\alpha = DL[S_1 op_1 p_1, \ldots; Q](\vec{t})$ be a dl-atom, then*

$$\tau^\Phi_{EH}(\alpha) = (\textstyle\bigwedge\{\phi \mid \phi \in \Phi\} \wedge \bigwedge_i \tau_{EH}(S_i op_i p_i) \supset Q(\vec{t})),$$
$$\tau_{EH}(\alpha) = (\textstyle\bigwedge_i \tau_{EH}(S_i op_i p_i) \supset Q(\vec{t})),$$

where
$$\tau_{EH}(S_i \uplus p_i) = (\forall \vec{x}.p_i(\vec{x}) \supset S_i(\vec{x})),$$
$$\tau_{EH}(S_i \cup\!\!\!\!-\, p_i) = (\forall \vec{x}.p_i(\vec{x}) \supset \neg S_i(\vec{x})), \textit{ and}$$
$$\tau_{EH}(S_i \Cap p_i) = (\forall \vec{x}.\neg \mathsf{L} p_i(\vec{x}) \supset \neg S_i(\vec{x})).$$

Let $\tau \in \{\tau_{EH}, \tau^\Phi_{EH}\}$. Then, $\tau(l) = l$, for each literal l, and

$$\tau(r) = (\forall)\ [\textstyle\bigwedge_i \mathsf{L}\tau(b_i) \wedge \bigwedge_{b_i \notin DL(r)} b_i \wedge \bigwedge_j \neg \mathsf{L} c_j] \supset \bigvee_k (h_k \wedge \mathsf{L} h_k),$$

for each r in P. Furthermore,

$$\tau_{EH}(\Phi, P) = \Phi \cup \{\tau_{EH}(r) \mid r \in P\},$$
$$\tau^\Phi_{EH}(P) = \{\tau^\Phi_{EH}(r) \mid r \in P\}.$$

Notice that rules without dl-atoms are embedded as before. Furthermore, dl-atoms are embedded as implications, where the antecedent consists of "hypothetical" additions of the extensions of rules predicates to classical predicates. Consider the DL atom $\alpha = DL[S \uplus p; Q](t)$; the embedding $\tau^\Phi_{EH}(\alpha) = (\bigwedge \Phi \wedge (\forall \vec{x}.p(\vec{x}) \supset S(\vec{x})) \supset Q(t))$ can be read as "whenever the extension of p is a subset of the extension of S, $Q(t)$ follows from Φ". This is analogous to the semantics of dl-atoms presented above.

Recall that a Herbrand interpretation I is a set of ground literals, and thus an FO theory. Recall also that $I^\approx$ is obtained from I by replacing every occurrence of '$=$' with '$\approx$'. The following theorem establishes a one-to-one correspondence between weak answer sets of dl-programs and consistent stable expansions of their embeddings.

Theorem 4. *A Herbrand interpretation I is a weak answer set of a dl-program (Φ, P) iff there is a consistent stable expansion T of $\tau^\Phi_{EH}(P)^\approx$ such that $I^\approx = T \cap \mathcal{L}^P_{gl}$. Moreover, if Φ is consistent and equality-free, there is a consistent stable expansion T' of $\tau_{EH}(\Phi, P)$ such that $I = T' \cap \mathcal{L}^P_{gl}$.*

To prove Theorem 4, given a dl-program (Φ, P) and a Herbrand interpretation I, we first "evaluate" all dl-atoms $\alpha \in DL(gr(P))$, by treating them as atoms and adding the ground atoms that are true to I. Let

$$I^{P,\Phi} = I \cup \{\alpha \in DL(gr(P)) | I \text{ is a model of } \alpha \text{ under } \Phi\},$$
$$P^{I,\Phi} = P \cup (I^{P,\Phi} \setminus I).$$

We obtain the following lemma.

Lemma 5. *Let (Φ, P) be a dl-program and I a Herbrand interpretation. Then, I is a weak answer set of (Φ, P) iff $I^{P,\Phi}$ is a stable model of $P^{I,\Phi}$.*

Proof. (*Sketch*) $P^{I,\Phi}$ is an "evaluation" of the dl-atoms in the spirit of the weak dl-transform and there is a one-to-one correspondence between the the reduct of $P^{I,\Phi}$ with respect to I and the weak dl-transform of P. □

Proof of Theorem 4. (*Sketch*) One can show that satisfaction of a ground dl-atom $\alpha = DL[\ldots; Q](\vec{t})$ in I under a theory Φ can be reduced to checking autoepistemic entailment:

$$I \text{ is a model of } \alpha \text{ under } \Phi \text{ iff } I^\approx \models_{I^\approx} \tau^\Phi_{EH}(\alpha)^\approx. \quad (5)$$

($\Rightarrow$) Let I be a weak answer set of (Φ, P). Consider $\tau_{EH}(P^{I,\Phi})^\approx$. By Lemma 5, $I^{P,\Phi}$ is a stable model of $P^{I,\Phi}$. So, by Proposition 3 and the definition of $I^{P,\Phi}$, $\tau_{EH}(P^{I,\Phi})^\approx$ has a consistent stable expansion T such that $I^\approx = T \cap \mathcal{L}^P_{gl}$.

Consider now a ground dl-atom α. We have that I is a model of α under Φ iff $\alpha \in I^{P,\Phi}$. Since $I^{P,\Phi}$ is a stable model of $P^{I,\Phi}$, we can conclude that I is a model of α under Φ iff $\alpha \in T_{ogl}$, by Proposition 3, and $\tau_{EH}(P^{I,\Phi})^\approx \models_{T_{ogl}} \alpha$, by Proposition 1.

Replacing α suitably with $\mathsf{L}\tau^\Phi_{EH}(\alpha)^\approx$ in T_{ogl}, we obtain T'_{ogl}; exploiting (5) and the shape of $\tau_{EH}(P^{I,\Phi})^\approx$, one can show that $T' = \{\phi \mid \tau^\Phi_{EH}(P)^\approx \models_{T'_{ogl}} \phi\}$ is a stable expansion of $\tau^\Phi_{EH}(P)^\approx$ such that $T \cap \mathcal{L}^P_{gl} = T' \cap \mathcal{L}^P_{gl}$.

The second part of the theorem can be shown using similar considerations, with the realization that there is no interaction between Φ and the embedding of the rules, since no predicates in Φ occur in rule heads and Φ is equality-free.

($\Leftarrow$) Let T be a consistent stable expansion of $\tau^\Phi_{EH}(P)^\approx$. Consider $\tau^\Phi_{EH}(P)'$, which is obtained from $\tau^\Phi_{EH}(P)^\approx$ by replacing every $\neg\mathsf{L}\tau^\Phi_{EH}(\alpha)^\approx$ with $\neg\mathsf{L}\alpha^\approx$ and $\mathsf{L}\tau^\Phi_{EH}(\alpha)^\approx$ with $(\alpha \wedge \mathsf{L}\alpha)^\approx$, where α is a dl-atom (viewed as predicate). Let

$$\tau^\Phi_{EH}(P)'' = \tau^\Phi_{EH}(P)' \cup \{(\alpha \wedge \mathsf{L}\alpha)^\approx \mid \tau^\Phi_{EH}(\alpha)^\approx \in T\}.$$

Clearly, $\tau^\Phi_{EH}(\Phi, P)''$ has a consistent stable expansion T' such that $T \cap \mathcal{L}^P_{gl} = T' \cap \mathcal{L}^P_{gl}$.

We need to show that $I = T'^\approx \cap \mathcal{L}^P_{gl}$ is a weak answer set of (Φ, P). We first establish that $\tau^\Phi_{EH}(P)'' = \tau_{EH}(P^{I,\Phi})^\approx$: indeed, it is easy to verify that $I^\approx \models_{I^\approx} \tau^\Phi_{EH}(\alpha)^\approx$ iff

$\tau^{\Phi}_{EH}(\alpha)^{\approx} \in T$, for any ground dl-atom α. Therefore, $\tau^{\Phi}_{EH}(P)'' = \tau_{EH}(P^{I,\Phi})^{\approx}$ follows immediately from (5) and the definition of $\tau^{\Phi\approx}_{EH}(P)''$. Since T' is a consistent stable expansion of $\tau_{EH}(P^{I,\Phi})^{\approx}$, we can conclude by Lemma 5 and Proposition 3 that I is a weak answer set of (Φ, P).

The second part of the theorem follows analogously. □

Example 7. *Reconsider the dl-program (Φ, P) from Example 5. Since $\Phi = \emptyset$, the embeddings τ_{EH} and τ^{Φ}_{EH} are clearly equivalent. We consider τ_{EH}. Define $\alpha_1 = DL[A \uplus a; A](t)$ and $\alpha_2(x) = DL[A \uplus a; A](x)$. Then, $\tau_{EH}(\Phi, P)$ contains*

$$a(t) \wedge \mathsf{L}a(t),$$
$$\mathsf{L}\beta_1 \supset (b(t) \wedge \mathsf{L}b(t)) \vee (b'(t) \wedge \mathsf{L}b'(t)),$$
$$\forall x.b(x) \wedge \mathsf{L}b(x) \wedge \neg\mathsf{L}\beta_2(x) \supset c(x) \wedge \mathsf{L}c(x),$$

where

$$\beta_1 = \tau_{EH}(\alpha_1) = ((\forall x.a(x) \supset A(x)) \supset A(t)),$$
$$\beta_2(x) = \tau_{EH}(\alpha_2(x)) = ((\forall y.a(y) \supset \neg A(y)) \supset A(x)).$$

This clause set has two stable expansions, containing $\{a(t), b(t), c(t)\}$ and $\{a(t), b'(t)\}$, respectively, as expected; recall that (Φ, P) has the weak answer sets $\{a(t), b(t), c(t)\}$ and $\{a(t), b'(t)\}$.

Indeed, $a(t)$ is justified by the first clause: from $a(t)$ we can prove $\forall x.(a(x) \supset A(x)) \supset A(t)$, and thus β_1 and $\mathsf{L}\beta_1$ are in the expansion. By the second clause, we have either $b(t) \wedge \mathsf{L}b(t)$ or $b'(t) \wedge \mathsf{L}b'(t)$ in the expansion, but not both.

Consider the expansion with $b(t)$. For $x = t$, we cannot prove $\beta_2(x)$; indeed, from $\forall y.(a(y) \supset \neg A(y)) \supset A(t)$ and $a(t)$, we cannot infer $A(t)$. Hence, $\neg\mathsf{L}\beta_2(t)$ is in the expansion. From the third clause, we obtain $c(t)$ and $\mathsf{L}c(t)$; it is not applicable for $x = t' \neq t$, as $b(t')$ is not derivable.

If $b'(t)$ is in the expansion, the third clause can not be applied, so no ground atom involving c is included.

Observe that $\tau_{EH}(\Phi, P)$ is modular, whereas $\tau^{\Phi}_{EH}(P)$ is "semi-modular": the embedding of individual rules depends on Φ, but not on other rules.

The embedding of dl-programs might have also been defined as an extension of the HP embedding, provided that no strong negation is used in the rules: in the embedding of rules in Definition 2, the L operator would not be applied to positive body or head literals. The proof of faithfulness of this embedding is simply obtained from the proof of Theorem 4 by replacing every 'EH' with 'HP'.

Note that the embedding τ_{EH} is not faithful if equality is used or if Φ is inconsistent. Consider (Φ, P), with $\Phi = \{a = b\}$ and $P = \{p(a)\}$. This (Φ, P) has a single weak answer set $I = \{p(a)\}$, whereas $\tau_{EH}(\Phi, P)$ has a single consistent stable expansion T such that $T \cap \mathcal{L}_{ga} = \{p(a), p(b)\}$.

Considerations about the Strong Answer-Set Semantics In contrast to weak answer sets, in *strong answer sets* for dl-programs (Eiter *et al.* 2004) the dl-atoms are also subject to a minimality condition; conclusions about dl-atoms need to be "grounded" in the existing knowledge. In AEL, stable expansions may be "ungrounded", suggesting that there is no straightforward (modular) embedding of the strong answer set semantics into FO-AEL. Unsurprisingly, the above embedding for weak answer set semantics does not work for strong answer-set semantics in general.

We note here that there has been work on defining grounded stable expansions (e.g., by Marek & Truszczyński (1993)) and constructive semantics for autoepistemic logic (Denecker, Marek, & Truszczynski 2003), which are close to default logic. Such modified semantics might be suitable for capturing the strong answer set semantics in the modular fashion. However, as for this paper, we concentrate on standard first-order autoepistemic logic.

There is an embedding of the strong answer set semantics (not considering $\cap\!\!\!\!\!-$) into Lifschitz's (1991) logic MKNF,[1] and also a non-modular embedding of MKNF into FO-AEL, which is discussed later in this paper; the fragment of MKNF we consider includes the embedding of dl-programs. Therefore, by composing these embeddings one obtains an embedding of the strong answer set semantics into FO-AEL.

$\mathcal{DL}$+*log* and R-Hybrid Knowledge Bases

R-hybrid knowledge bases (Rosati 2005) (sometimes also referred to as $\mathcal{DL}+log$ (Rosati 2006)) are combinations of first-order theories and logic programs, interpreted under the stable-model semantics. We consider an extended version in which classical predicates may occur in negated body literals in rules, as considered by de Bruijn *et al.* (2007b).

An *extended r-hybrid knowledge base* is a pair $\mathcal{B} = (\Phi, P)$, where Φ is a finite FO theory over the SNA signature $\Sigma_\Phi = \langle \mathcal{F}, \mathcal{P}_\Phi \rangle$ and P is a safe logic program without strong negation over the signature $\Sigma = \langle \mathcal{F}, \mathcal{P}_\Phi \cup \mathcal{P}_P \rangle$, with $\mathcal{P}_\Phi \cap \mathcal{P}_P = \emptyset$, where both Φ and P are equality- and function-free. Moreover, we define $\Sigma_P = \langle \mathcal{F}, \mathcal{P}_P \rangle$. We call atoms over Σ_P *rule atoms* and atoms over Σ_Φ *classical atoms*. As usual, an interpretation $\mathcal{I}$ satisfies a negated atom $not\ a_i$ if $\mathcal{I} \not\models a_i$. $\mathcal{B}$ is an *r-hybrid knowledge base* (Rosati 2005) if P does not contain negated classical atoms.

If $\mathcal{I}$ is an interpretation of the language $\mathcal{L}$ with signature Σ and $\mathcal{L}^\Phi \subseteq \mathcal{L}$ has signature Σ_Φ, then $\mathcal{I}|_\Phi$ denotes the restriction of $\mathcal{I}$ to $\mathcal{L}^\Phi$. Given an interpretation $\mathcal{I}$ of $\mathcal{L}^\Phi$ and a ground logic program P over $\mathcal{L}$, the *projection of P with respect to $\mathcal{I}$*, denoted $\Pi(P, \mathcal{I})$, results from P by deleting

1. each rule $r \in P$ that has a classical atom A in the head such that $\mathcal{I} \models A$ or a negated classical atom A in the body such that $\mathcal{I} \not\models A$, and
2. from each remaining rule $r \in P$ each classical atom A in the head such that $\mathcal{I} \not\models A$ or a negated classical atom A in the body such that $\mathcal{I} \models A$.

Intuitively, $\Pi(P, \mathcal{I})$ "evaluates" P with respect to $\mathcal{I}$.

Given an r-hybrid KB $\mathcal{B} = (\Phi, P)$ over the language $\mathcal{L}$, an SNA interpretation $\mathcal{I}$ of $\mathcal{L}$ is an *NM-model* of $\mathcal{B}$ if $\mathcal{I}|_\Phi$ is a model of Φ and $\mathcal{I}|_P$ is a stable model of $\Pi(gr(P), \mathcal{I}|_\Phi)$.

Embedding r-Hybrid Knowledge Bases One might expect that the semantics of r-hybrid KBs corresponds to that of a straightforward combination $\Phi \cup \tau_{HP}(P)$, since in $\tau_{HP}(P)$ positive body atoms do not occur under the modal operator L. However, it turns out that this embedding is not faithful, because of the different interpretation of classical and rules predicates in r-hybrid KBs.

[1] Personal communication with Boris Motik.

Example 8. *Consider* $\mathcal{B} = (\Phi, P)$*, with* $\Phi = \{C(a)\}$ *and* $P = \{r(b) \leftarrow C(b);\, s \leftarrow not\ r(b)\}$*, where* C *is a classical predicate and* s, r *are rules predicates.* $\mathcal{B}$ *does not entail* s*: consider* $\mathcal{I} = \{C(a),\, C(b),\, r(b)\}$*. Clearly,* $\mathcal{I}|_\Phi \models \Phi$ *and* $\mathcal{I}|_P$ *is a stable model of* $\Pi(P, \mathcal{I}|_\Phi) = \{r(b); s \leftarrow not\ r(b)\}$*. Thus,* $\mathcal{I}$ *is an NM-model of* $\mathcal{B}$*. However, as easily seen,* s *is in each stable expansion of* $\Phi \cup \tau_{HP}(P)$*.*

An alternative embedding one might expect to work is inspired by the embedding of r-hybrid KBs into MKNF by Motik & Rosati (2007b). Rules of the form (3) would then be embedded as follows:

$$\left(\bigwedge_g (b_g) \wedge \bigwedge_h (\mathsf{L}b_h \wedge b_h) \wedge \bigwedge_i \neg \mathsf{L}c_i\right) \supset \left(\bigvee_j h_j \vee \bigvee_k (\mathsf{L}h_k \wedge h_k)\right),$$

where b_g, h_j are classical atoms and b_h, c_i, h_k are rule atoms. As shown next, also this embedding is not faithful.

Example 9. *The r-hybrid KB* $\mathcal{B}$ *of Example 8 is embedded as* $\Psi = \{C(a),\, C(b) \supset (\mathsf{L}r(b) \wedge r(b)),\, \neg \mathsf{L}r(b) \supset (\mathsf{L}s \wedge s)\}$*. Now, every stable expansion* T *of* Ψ *must contain* s*: Suppose* $s \notin T$*. By the third axiom in* Ψ*,* $r(b)$ *must be in* T*. Hence,* $\Psi \models_T r(b)$ *must hold. Consider the interpretation* $\mathcal{I} = \{C(a)\}$*. Clearly,* $\mathcal{I} \models_T \Psi$ *and* $\mathcal{I} \not\models_T r(b)$*. Thus,* $\Psi \not\models_T r(b)$*, which is a contradiction.*

We now consider an embedding of r-hybrid KBs that is based on the observation that classical predicates are interpreted classically. To this end, we introduce the set CIA_{Σ_Φ} of *classical interpretation* (CI) *axioms*, consisting of

CIA $\quad (\forall)\ \mathsf{L}P(\vec{x}) \equiv P(\vec{x}), \quad$ for each $P \in \mathcal{P}_\Phi$.

The embedding is the straightforward combination (union) of the FO theory, the embedding of the program, and the CI axioms. The following theorem establishes a one-to-one correspondence between NM-models of r-hybrid knowledge bases and consistent stable expansions of their embeddings, which implies correspondence between notions of consistency (existence of NM-model vs. existence of consistent stable expansion) and consequence (entailment vs. autoepistemic consequence).

Theorem 6. *Let* $\mathcal{I}$ *be an SNA interpretation and let* (Φ, P) *be an extended r-hybrid knowledge base. Then,* $\mathcal{I}$ *is an NM-model of* (Φ, P) *iff* $\tau_{EH}(P) \cup \Phi \cup CIA_{\Sigma_\Phi}$ *has a consistent stable expansion* T *such that* $T_{oga} = \{\alpha \in \mathcal{L}_{ga} \mid \mathcal{I} \models \alpha\}$*.*

Proof. Let ι be short for $\tau_{EH}(P) \cup \Phi \cup CIA_{\Sigma_\Phi}$, let $\mathcal{L}^\Phi$ and $\mathcal{L}^P$ denote the languages with the signatures $\langle \mathcal{F}, \mathcal{P}_\Phi \rangle$ and $\langle \mathcal{F}, \mathcal{P}_P \rangle$, respectively, and let Φ denote the language with the signature $\langle \mathcal{F}, \mathcal{P}_\Phi \cup \mathcal{P}_P \rangle$.

($\Rightarrow$) Assume $\mathcal{I}$ is an NM-model of $\mathcal{B}$. We define $T_{oga} = \{\alpha \in \mathcal{L}_{ga} \mid \mathcal{I} \models \alpha\}$. We will show that

$$T_{oga} = \{\alpha \in \mathcal{L}_{ga} \mid \iota \models_{T_{oga}} \alpha\}. \quad (6)$$

It follows by Proposition 1 that there is a stable expansion T of ι and that T is the unique expansion such that $T_{oga} = T \cap \mathcal{L}_{ga}$.

Consider a ground atom $\alpha \in T_{oga}$. If $\alpha \in \mathcal{L}^\Phi$, then $\iota \models_{T_{oga}} \alpha$, by the CI axioms.

Since $\mathcal{I}$ is an NM-model of (Φ, P), it is easy to see that

$$\{\alpha \in \mathcal{L}^P \mid \tau_{EH}(\Pi(gr(P), \mathcal{I}|_\Phi)) \models_{T_{oga}} \alpha\} = \{\alpha \in \mathcal{L}^P \mid \tau_{EH}(gr(P)) \cup \Phi \cup CIA_{\Sigma_\Phi} \models_{T_{oga}} \alpha\}. \quad (7)$$

By the standard-names assumption we can conclude

$$\{\alpha \in \mathcal{L}^P \mid \tau_{EH}(gr(P)) \cup \Phi \cup CIA_{\Sigma_\Phi} \models_{T_{oga}} \alpha\} = \{\alpha \in \mathcal{L}^P \mid \iota \models_{T_{oga}} \alpha\}. \quad (8)$$

As $\mathcal{I}|_P$ is a stable model of $\Pi(gr(P), \mathcal{I}|_\Phi)$, we know by construction of T_{oga} and the results in Propositions 1 and 3 that $T' = \{\phi \mid \tau_{EH}(\Pi(gr(P), \mathcal{I}|_\Phi)) \models_{T_{oga}} \phi\}$ is a stable expansion of $\tau_{EH}(\Pi(gr(P), \mathcal{I}|_\Phi))$ and $T'_{oga} = T_{oga}$. Combined with (7) and (8), we can conclude that $\alpha \in \mathcal{L}^P$ implies $\iota \models_{T_{oga}} \alpha$.

To show the converse, consider a ground atom $\alpha \in \mathcal{L}$ such that $\iota \models_{T_{oga}} \alpha$. This means that ($*$) for every interpretation $\mathcal{I}'$ such that $\langle \mathcal{I}', T_{oga} \rangle$ is a model of ι, it holds that $\mathcal{I}' \models_{T_{oga}} \alpha$. We show that $\mathcal{I} \models_{T_{oga}} \iota$.

$\langle \mathcal{I}, T_{oga} \rangle$ is clearly a model of Φ. $\langle \mathcal{I}, T_{oga} \rangle$ is a model of $\tau_{EH}(\Pi(gr(P), \mathcal{I}|_\Phi))$ by the fact that $\mathcal{I}|_P$ is a model of $\Pi(gr(P), \mathcal{I}|_\Phi)$ and since $\alpha \notin T_{oga}$ iff $\mathcal{I} \not\models \alpha$, for any $\alpha \in \mathcal{L}_{ga}$. Then, by construction of $\Pi(gr(P), \mathcal{I}|_\Phi)$, $\langle \mathcal{I}, T_{oga} \rangle$ must be a model of $\tau_{EH}(gr(P))$ and, by the SNA, of $\tau_{EH}(P)$. Finally, $\langle \mathcal{I}, T_{oga} \rangle$ is a model of CIA_{Σ_Φ} since there is no $\alpha \in T_{oga}$ such that $\mathcal{I} \not\models \alpha$. Therefore, $\mathcal{I} \models_{T_{oga}} \iota$.

If $\alpha \notin T_{oga}$, then $\mathcal{I} \not\models_{T_{oga}} \alpha$ (by construction of T_{oga}), contradicting ($*$). This establishes (6). Since $\langle \mathcal{I}, T_{oga} \rangle$ is a model of ι, T must be consistent.

($\Leftarrow$) Assume ι has a consistent stable expansion T. Let $\mathcal{I} = \langle U, \cdot^I \rangle$ be an interpretation such that for every atom $\alpha \in \mathcal{L}$, $\mathcal{I} \models \alpha$ iff $\alpha \in T_{oga}$. $\mathcal{I}|_\Phi \models \Phi$ must hold because, for any atom $\alpha \in \mathcal{L}^\Phi$ and any model $\langle \mathcal{I}', T_{oga} \rangle$ of ι, $\mathcal{I}' \models_{T_{oga}} \alpha$ iff $\alpha \in T_{oga}$, by the CIA axioms. It remains to be shown that $\mathcal{I}|_P$ is a stable model of $\Pi(gr(P), \mathcal{I}|_\Phi)$.

Clearly, by the SNA, equation (8) holds. It is also easy to verify that (7) holds, because of the correspondence between T_{oga} and $\mathcal{I}$. Therefore, $\tau_{EH}(\Pi(gr(P), \mathcal{I}|_\Phi))$ has a stable expansion T' such that $T'_{oga} = T_{oga} \cap \mathcal{L}^P$. By Proposition 3, T'_{oga}, and hence $\mathcal{I}|_P$, is a stable model of $\Pi(gr(P), \mathcal{I}|_\Phi)$. It follows that $\mathcal{I}$ is an NM-model of (Φ, P). $\square$

Observe that the result clearly extends to a variant of r-hybrid KBs with finite sets of standard names, when also considering SNA interpretations with finite domains in FO-AEL. Furthermore, the embedding is clearly modular: individual axioms in CIA_{Σ_Φ} depend only on the use of individual predicate symbols in the axioms in Φ.

One might expect that considering a weaker variant of the CI axioms, with implication instead of equivalence (i.e., $(\forall)\mathsf{L}P(\vec{x}) \supset P(\vec{x})$), would lead to a similar result. However, this is not the case. Consider, for example, the r-hybrid KB $(\{a \vee b\}, \{p \leftarrow a; p \leftarrow b\})$, which has three NM-models: in one a is satisfied, in one b, and in one both a and b; furthermore, in all three p is satisfied. With weakened CI axioms, the embedding has three stable expansions corresponding to the NM-models, but it also has one expansion containing neither a, b, nor c (satisfaction in some interpretation does not imply inclusion in the expansion).

Theorem 6 holds if we use the HP embedding instead of the EH embedding, again also for a variant with finite sets of standard names. The resultant embedding is signature-modular (but not modular). Weakening the CI axioms to implication leads again to a loss of faithfulness of the embedding in general, but is immaterial for the classical predicates that occur in the logic program; in the example above, we would obtain three stable expansions, as desired. Investigating embeddings of disjunction-free r-hybrid KBs with weakened CI axioms is left for future work.

In embedding r-hybrid KBs into FO-AEL we used the non-trivial embedding $\tau_{CIA}(\Phi) = \Phi \cup CIA_{\Sigma_\Phi}$ for a classical FO theory Φ into FO-AEL, rather than the trivial one with $\tau(\Phi) = \Phi$. Theorem 6 tells us that satisfiability and ground entailment for Φ correspond with the existence of a consistent stable expansion and autoepistemic consequence for $\tau_{CIA}(\Phi)$, respectively; the same holds for $\tau(\Phi)$. A notable difference is that any satisfiable Φ has a single (consistent) stable expansion, while $\tau_{CIA}(\Phi)$ typically has many stable expansions.

Relating MKNF and FO-AEL

The logic MKNF of *minimal knowledge and negation as failure* (Lifschitz 1991) is a non-monotonic logic that has been developed for formalizing the stable-model semantics. It has been slightly adapted and used as a unifying formalism for combining ontologies and logic programs by Motik & Rosati (2007a), who we follow in our treatment.

MKNF is a first-order modal logic with a knowledge operator K and negation-as-failure operator not. As in FO-AEL, there are no restrictions on the use of quantifiers and nesting of modal operators. We denote by $\mathcal{L}_{MKNF}$ the MKNF language obtained from a language $\mathcal{L}$, and by $\mathcal{L}^1_{MKNF}$ the set of MKNF formulas with no nesting of modal operators.

An *MKNF structure* is a triple $(\mathcal{I}, M, N)$, where $\mathcal{I}$ is an interpretation and M, N are non-empty sets of interpretations. Satisfaction of non-modal formulas is defined relative to $\mathcal{I}$ in the usual way. Satisfaction of modal formulas is defined as follows:

$$(\mathcal{I}, M, N) \models^{\mathsf{K},\mathsf{not}} \mathsf{K}\phi \quad \text{iff } \forall \mathcal{I}' \in M, (\mathcal{I}', M, N) \models^{\mathsf{K},\mathsf{not}} \phi;$$
$$(\mathcal{I}, M, N) \models^{\mathsf{K},\mathsf{not}} \mathsf{not}\,\phi \text{ iff } \exists \mathcal{I}' \in N, (\mathcal{I}', M, N) \not\models^{\mathsf{K},\mathsf{not}} \phi.$$

This extends to arbitrary formulas in the usual way.

A non-empty set of SNA interpretations M satisfies a closed MKNF formula ϕ, denoted $M \models^{\mathsf{K},\mathsf{not}} \phi$, if $(\mathcal{I}, M, M) \models^{\mathsf{K},\mathsf{not}} \phi$ for each $\mathcal{I} \in M$. M is an *MKNF model* of ϕ if $M \models^{\mathsf{K},\mathsf{not}} \phi$ and there is no MKNF structure $(\mathcal{I}', M', M)$ with $\mathcal{I}' \in M'$ such that $M' \supsetneq M$ and $(\mathcal{I}', M', M) \models^{\mathsf{K},\mathsf{not}} \phi$.

Hybrid MKNF Knowledge Bases In their extension of ontologies with MKNF rules, Motik & Rosati (2007a; 2007b) consider MKNF formulas of a specific form, called *hybrid MKNF$^+$ knowledge bases* (KBs):

$$\varphi \quad = \quad \mathsf{K}(\textstyle\bigwedge \Phi) \wedge \bigwedge \Pi, \tag{9}$$

where $\Phi \subseteq \mathcal{L}$ is a finite FO theory and Π is a finite set of *MKNF rules*, which are of the form

$$(\forall)\ b_1 \wedge \cdots \wedge b_n \wedge \mathsf{K}c_1 \wedge \cdots \wedge \mathsf{K}c_o \wedge \mathsf{not}\,d_1 \wedge \cdots \wedge \mathsf{not}\,d_p \supset g_1 \vee \cdots \vee g_q \vee \mathsf{K}h_1 \vee \cdots \vee \mathsf{K}h_r, \tag{10}$$

where b_i, c_j, d_k, g_l, h_m are literals in $\mathcal{L}$; if $n = o = p = 0$, the symbol $\supset$ is dropped; the empty head ($q = r = 0$) is written as $\perp$. A *hybrid MKNF knowledge base* is a hybrid MKNF$^+$ KB in which each MKNF rule (10) has $n = q = 0$.

The structure of the remainder of this section is as follows. We first consider the embedding of MKNF clauses into FO-AEL, which serves as a stepping stone towards embedding hybrid MKNF KBs. We then show that there is no modular embedding of such MKNF KBs into FO-AEL and present a non-modular embedding of generalized MKNF clauses, which generalize hybrid MKNF KBs.

Embedding MKNF Clauses *Protected literals* are modal atoms of the forms $\mathsf{K}\phi$ and $\mathsf{not}\,\phi$, where ϕ is an objective literal. *PL-theories* are MKNF theories consisting only of propositional combinations of protected literals. As independently shown by Lifschitz & Schwarz (1993) and Chen (1993), there is a faithful modular embedding $\tau_{AEL'}(\Phi)$ of (propositional) PL-theories Φ into AEL.[2] We lift this result to MKNF clause sets.

An *MKNF clause* is a formula

$$Q_1 x_1 \ldots Q_n x_n.\ A_1 \vee \cdots \vee A_m \vee \neg B_1 \vee \cdots \vee \neg B_l, \tag{11}$$

where all Q_i are quantifiers and A_j, B_k are protected literals. Observe that an MKNF rule of form (10) is an MKNF clause if all atoms are modal, i.e., if $n = q = 0$.

For $\Phi \subseteq \mathcal{L}^1_{MKNF}$, let $\tau_{AEL'}(\Phi) = \{\tau_{AEL'}(\phi) \mid \phi \in \Phi\}$, where $\tau_{AEL'}(\phi)$ results from ϕ by replacing each $\mathsf{K}\psi$ in ϕ with $(\psi \wedge \mathsf{L}\psi)$ and each $\mathsf{not}\,\psi$ in ϕ with $\neg\mathsf{L}\psi$.

Proposition 7. *Let $\Phi \subseteq \mathcal{L}_{MKNF}$ be a set of MKNF clauses. A non-empty set M of interpretations is an MKNF model of $\Phi^{\approx}$ iff M is an AEL model of $\tau_{AEL'}(\Phi^{\approx})$.*

Proof. (*Sketch*) The proof is an extension of the one for the propositional case with MBNF by Lifschitz & Schwarz (1993). MBNF interpretations are of the form $\langle \mathcal{I}, M \rangle$. Satisfaction and models for MBNF are defined like for MKNF, with the exception that $\mathcal{I}$ is not required to be included in M. Since all objective formulas in Φ occur in the scope of a modal operator, $\mathcal{I}$ does not play a role, so satisfaction and models in MBNF and in MKNF are equivalent.

The proof by Lifschitz & Schwarz (1993) uses *reflexive* autoepistemic logic, whose first-order variant is obtained from our model-based definition of AEL by modifying satisfaction of modal formulas $\mathsf{L}\phi$ in (2) to additionally require $\langle \mathcal{I}, M \rangle, B \models \phi$.

The proof of the main theorem of Lifschitz & Schwarz (1993) extends to the first-order setting with SNA as follows: an SNA interpretation $\mathcal{I}$ can be viewed as a propositional interpretation that consists of all ground atoms satisfied by $\mathcal{I}$; the structure of the proof remains the same. To see why the main lemma and Lemmas 4.1 and 4.2 by Lifschitz & Schwarz (1993) extend to MKNF clauses, it is sufficient to realize that $(\mathcal{I}, M, N) \models^{\mathsf{K},\mathsf{not}} \forall x.\phi(x)$ iff $(\mathcal{I}, M, N) \models^{\mathsf{K},\mathsf{not}} \phi(t)$, for every $t \in \mathcal{N}$; similar for $\exists x.\phi(x)$. $\square$

For a logic program P, let $\tau_{MKNF}(P)$ consist of

$$(\forall)\ \mathsf{K}b_1 \wedge \cdots \wedge \mathsf{K}b_m \wedge \mathsf{not}\,c_1 \wedge \cdots \wedge \mathsf{not}\,c_n \supset \mathsf{K}h_1 \vee \cdots \vee \mathsf{K}h_l,$$

[2] Actually, they all considered *MBNF*, which is equivalent to MKNF if all objective atoms occur in the scope of a modal operator.

for each rule of form (3) in P, following Lifschitz (1991). Clearly, $\tau_{AEL'}(\tau_{MKNF}(P)) = \tau_{EH}(P)$. The following corollary, which establishes a correspondence between the embedding into MKNF and the EH embedding, follows immediately from this equation and Proposition 7.

Corollary 8. *Let P be a logic program and M a non-empty set of interpretations. Then, M is an MKNF model of $\tau_{MKNF}(P)^{\approx}$ iff M is an AEL model of $\tau_{EH}(P)^{\approx}$.*

This result does not carry over to the HP embedding, even when considering only safe rules, because of the PI axioms in the HP embedding. Note that the treatment of equality in the translation is not an issue, since the variant of MKNF in Motik & Rosati (2007a) treats equality like we do.

It turns out that when adding an ontology to a set of MKNF rules, there is not always a modular embedding into FO-AEL, as shown below.

Considering Hybrid MKNF Knowledge Bases Gottlob (1995) showed that there is no faithful modular embedding of default logic (Reiter 1987) into AEL. Since default logic has a modular embedding into MKNF, there is no modular embedding of MKNF into AEL in the general case. We adapt Gottlob's proof to our notion of modularity and to atomic defaults, whose embedding into MKNF yields a hybrid MKNF KB.

We consider variable-free default theories $\langle \Phi, D \rangle$, where $\Phi \subseteq \mathcal{L}$ is a background theory and D is a set of *atomic defaults*, which are of the form $\alpha : M\beta_1, \ldots, M\beta_n \,/\, \omega$, where α, $\beta_1, \ldots, \beta_n$, and ω are atoms.

Chen (1994) showed that there is a faithful modular embedding into AEL if, additionally, Φ is a set of conjunctions of literals. By an adaptation of Gottlob (1995, Theorem 3.2) we show that this result does not extend to arbitrary Φ. Note that Gottlob assumed the background theory Φ to remain unchanged in the embedding; we do not make this assumption.

Proposition 9. *There is no faithful modular embedding of default logic into standard AEL, even for atomic defaults.*

Proof. (*Sketch*) Gottlob (1995) uses three default theories in his proof, which all have different background theories,

$$W_0 = \emptyset,\ W_1 = \{a\},\ \text{and}\ W_2 = \{a \supset b\},$$

but share the same set of defaults

$$D = \Big\{ a \supset b : Ma \,/\, a,\ a : Mb \,/\, b \Big\}.$$

He used a different notion of modularity: the embedding of the background theory W and the set of defaults D must be independent, but not for individual defaults; moreover, W is added directly to the embedding of the defaults, while we allow W to be transformed as well. Gottlob's proof can be adapted to our setting by considering embeddings $tr(W) \neq W$. In essence, $tr(W_i)$ must amount to W_i, for each W_i, by faithfulness of tr: for our purposes, we may assume that $tr(W_i)$ consists of W_i and a set of clauses in AEL normal form that have a limited structure. This is due to the fact that for each objective set of formulas $W \supseteq W_i$, $W \cup tr(W_i)$ must have a (single) stable expansion whose objective part is $\{\phi \in \mathcal{L} \mid W_i \models \phi\}$. These clauses can then be removed.

Then, the defaults in D are not atomic, as one has $a \supset b$ as its prerequisite. We can replace $a \supset b$ with a new atom p, obtaining D', and add $p \equiv (a \supset b)$ to W_i, obtaining W_i', for $i \in \{1, 2, 3\}$. Clearly, (W_i, D) and (W_i', D') have isomorphic default extensions. By exploiting modularity of $tr(W)$, the proof can then be easily adapted to this setting. ☐

Following Lifschitz (1991), a default theory $\langle \Phi, D \rangle$ is embedded into MKNF by adding to Φ the MKNF rule

$$\mathsf{K}\alpha \wedge \mathsf{not}\,\neg\beta_1 \wedge \cdots \wedge \mathsf{not}\,\neg\beta_n \supset \mathsf{K}\omega,$$

for each default $\alpha : M\beta_1, \ldots, M\beta_n \,/\, \omega$ in D. Observe that if the defaults are atomic, the result is a hybrid MKNF KB of the form (9).

Proposition 10. *There is no faithful modular embedding of hybrid MKNF knowledge bases, and hence of hybrid MKNF$^+$ knowledge bases, into FO-AEL.*

Embedding Generalized MKNF Clauses Rosati (1997) gave a non-modular faithful embedding of MKNF into AEL in the propositional case, based on the embedding of default logic into AEL by Schwarz (1996). We adopt these techniques to embed *generalized MKNF clauses*, which are like MKNF clauses (11) except that A_i, B_j may be arbitrary modal atoms in $\mathcal{L}^1_{MKNF}$.

Let Φ be a set of k generalized MKNF clauses and let $p_1, \ldots, p_m$ be the list of all predicate symbols appearing in Φ. We fix k additional lists of new predicate symbols $p^i_1, \ldots, p^i_m$, $1 \leq i \leq k$. Let $\phi(p^i)$ denote the result of substituting $p^i_1, \ldots, p^i_m$ for the $p_1, \ldots, p_m$ in ϕ. The embedding $\tau_{AEL''}(\phi)$ is obtained by replacing in ϕ each $\mathsf{K}\psi$ with $(\psi \wedge \mathsf{L}\psi \wedge \bigwedge_{j=1}^{k} \psi(p^j))$ and each $\mathsf{not}\,\psi$ with $\neg\mathsf{L}\psi$. Note that the additional predicates p^j "block" ungrounded inferences; see Rosati (1997) and Schwarz (1996) for more details.

We define $\tau_{AEL}(\Phi) = \{\tau_{AEL}(\varphi) \mid \varphi \in \Phi\}$, where

$$\tau_{AEL}(\varphi) = Q_1x_1 \ldots Q_{n_i}x_{n_i}.\tau_{AEL'}(\textstyle\bigvee_g A_g \vee \bigvee_h B_h) \vee \\ \tau_{AEL''}(\textstyle\bigvee_g A_g \vee \bigvee_h B_h).$$

We then have the following model correspondence:

Proposition 11. *Let $\Phi \subseteq \mathcal{L}_{MKNF}$ be a finite set of generalized MKNF clauses and let M be a non-empty set of interpretations. Then, M is an MKNF model of $\Phi^{\approx}$ iff M is an AEL model of $\tau_{AEL}(\Phi^{\approx})$.*

The proof of this proposition is a straightforward extension of the proof of Rosati (1997, Theorem 8), along the same lines as the last paragraph in the proof of Proposition 7.

We can exploit this result for finite propositional MKNF theories, since each such theory can be rewritten to a conjunction of generalized MKNF clauses (Rosati 1997). However, this does not extend to the first-order case. For example, in $\mathsf{K}(\exists x.a(x) \vee \neg\mathsf{K}b(x))$, K does not distribute over extensional quantification (cf. Motik & Rosati (2007b)).

Nonetheless, hybrid MKNF$^+$ KBs (9) can be straightforwardly rewritten to generalized clauses, since in MKNF a conjunction of closed formulas $\phi_1 \wedge \cdots \wedge \phi_n$ is equivalent to the theory $\{\phi_1, \ldots, \phi_n\}$ and an MKNF rule (10) can be equivalently rewritten to a generalized MKNF clause (cf. Motik & Rosati (2007b)):

$$(\forall)\ \neg\mathsf{K}c_1 \vee \cdots \vee \neg\mathsf{K}c_o \vee \neg\mathsf{not}\, d_1 \vee \cdots \vee \neg\mathsf{not}\, d_p \vee$$
$$\mathsf{K}h_1 \vee \cdots \vee \mathsf{K}_r \vee \mathsf{K}(b_1 \wedge \cdots \wedge b_n \supset g_1 \vee \cdots \vee g_q).$$

As for arbitrary MKNF theories, it seems that they may be rewritten as sets of generalized MKNF clauses using auxiliary predicates. The resulting theory can then be rewritten as a FO-AEL theory using the embedding that we presented.

Standard vs. Any- and All-Name Semantics

The any- and all-name semantics for FO-AEL were defined by Konolige (1991). These semantics are obtained from the standard-names semantics by considering arbitrary (and not only SNA) signatures and interpretations and changing the definition of satisfaction of modal atoms. We denote satisfaction for the any-names semantics by $\models^{\mathrm{E}}$ and for the all-name semantics by $\models^{\mathrm{A}}$. The conditions for these relations for modal atoms are as follows (we consider here only the case of one free variable for simplicity):

$$\mathcal{I}, B \models^{\mathrm{E}}_{\Gamma} \mathsf{L}\phi(x) \text{ iff } \phi(t)\in\Gamma \text{ for some } t\in\mathcal{N} \text{ s.t. } t^I = x^B;$$
$$\mathcal{I}, B \models^{\mathrm{A}}_{\Gamma} \mathsf{L}\phi(x) \text{ iff } \phi(t)\in\Gamma \text{ for all } t\in\mathcal{N} \text{ s.t. } t^I = x^B.$$

An individual k is *unnamed* if there is no $t \in \mathcal{N}$ such that $t^I = k$. The notions of stable$^{\mathrm{E}}$ and stable$^{\mathrm{A}}$ expansions are defined analogous to the SNA case. The following example, partly due to (Levesque 1990), illustrates some of the differences between stable, stable$^{\mathrm{E}}$, and stable$^{\mathrm{A}}$ expansions.

Example 10. *All stable expansions include the converse Barcan formula*

$$\mathsf{L}(\forall x.\varphi(x)) \supset \forall x.\mathsf{L}\varphi(x)$$

for any FO formula φ. It is not included in all stable$^{\mathrm{E}}$, nor in all stable$^{\mathrm{A}}$ expansions, because in the absence of the SNA there might be an unnamed individual k in models of an extension of φ, which means that there is no ground term t representing k such that $\varphi(t)$ is included in the expansion.

Example 11. *Consider the theory $\Phi = \mathsf{L}p(a)$. $\Phi^{\approx}$ has one stable expansion T^s; Φ has one stable$^{\mathrm{E}}$ expansion T^E and one stable$^{\mathrm{A}}$ expansion T^A. Consider the formulas*

$$\phi = \forall x.x = a \supset \mathsf{L}p(x) \quad \textit{and}$$
$$\psi = \forall x.x = b \supset \neg\mathsf{L}p(x).$$

T^s includes both $\phi^{\approx}$ and $\psi^{\approx}$. T^E includes ϕ, but not ψ, because in a given interpretation a and b may be assigned to the same individual, and $p(a)$ is included in T^E, so this interpretation does not satisfy ψ. T^A, on the other hand, includes ψ, but not ϕ, by an analogous argument.

Stable expansions have a characterization purely in terms of standard entailment $\models_{\mathrm{FOL}}$, where modal atoms $\mathsf{L}\phi$ with free variables $x_1, \ldots, x_n$ are viewed as n-ary predicates.

Proposition 12. *Let $\Phi \subseteq \mathcal{L}_{\mathsf{L}}$ be a theory and let $T \subseteq \mathcal{L}_{\mathsf{L}}$ be a belief set. Then, $T^{\approx}$ is a stable expansion of $\Phi^{\approx}$ iff $T = \{\phi \mid \Phi \cup \{\mathsf{L}\psi \mid \psi \in T\} \cup \{\neg\mathsf{L}\psi \mid \psi \notin T\} \models_{\mathrm{FOL}} \phi\}$.*

Proof. (*Sketch*) We prove the proposition for the case of $\Phi^{\approx}$ and FOL entailment under the SNA; the proposition then follows from the classical results about reasoning with FOL with equality. We omit the $^{\approx}$ modifier and show that $\Phi^{\cup T} \models_{FOL} \phi$ iff $\Phi \models_T \phi$, where $\Phi^{\cup T} = \{\phi \mid \Phi \cup \{\mathsf{L}\psi \mid \psi \in T\} \cup \{\neg\mathsf{L}\,\psi \mid \psi \notin T\}$; the correspondence of the definitions is then immediate.

Consider an interpretation $\mathcal{I}$ such that $\mathcal{I} \models_T \Phi$ and let $\mathcal{I}'$ be an interpretation such that $\mathcal{I}'|_{\mathcal{L}} = \mathcal{I}$ and $\mathcal{I}' \models_{FOL} \mathsf{L}\psi$ iff $\psi \in T$. Clearly, $\mathcal{I}' \models_{FOL} \Phi^{\cup T}$ and $\mathcal{I} \models_T \phi$ iff $\mathcal{I}' \models_{FOL} \phi$. It follows that $\Phi \models_T \phi$ implies $\Phi^{\cup T} \models_{FOL} \phi$. The converse is similar. □

This fails for the any- and all-name semantics: consider the theory $\Phi = \{\, p(a), \exists x.x \neq a, \forall x.\neg\mathsf{L}p(x) \supset q \,\}$ in a language with only one name, a. Then, Φ has a stable$^{\mathrm{E}}$ expansion T with $T_{oga} = \{p(a), a = a, q\}$: each model must have at least two individuals, one of which is unnamed, due to the second axiom in Φ; analogous for the all-name semantics. Now, $\Phi \cup \{\mathsf{L}\psi \mid \psi \in T\} \cup \{\neg\mathsf{L}\psi \mid \psi \notin T\} \not\models_{\mathrm{FOL}} q$: the single ground instance of $p(x)$ is $p(a)$, which belongs to T, so there is no $p(t)$ such that $p(t) \notin T$, with $t \in \mathcal{N}$.

When considering the any- or all-name semantics, default uniqueness of names needs to be axiomatized when embedding logic programs, following de Bruijn *et al.* (2007). Let UNA_{Σ} denote the set of *unique-names assumption* (UNA) *axioms*

UNA $\quad \neg\mathsf{L}(t_1 = t_2) \supset t_1 \neq t_2, \quad$ for distinct names t_1, t_2.

When the UNA axioms are included, the results on embedding dl-programs into FO-AEL extend to the any- and all-name semantics provided that Φ is equality-free and the embedding of $S_i \Cap p_i$ additionally includes $\bigwedge_j \mathsf{L}x_j = x_j$ in the antecedent, in order to ensure the implication does not become “active” when quantifying over unnamed individuals. For arbitrary theories this is no longer true, because the UNA axioms may interact with equality in Φ.

For r-hybrid KBs, our results do not extend to the any- or all-name semantics for finite signatures, simply because unnamed individuals might compromise the embedding (consider, e.g., $\mathcal{B} = (\{C(a), \exists x.\neg C(x)\}, \emptyset)$ with standard names $\{a\}$). The same holds if implicitly for SNA interpretations infinitely many standard names were added.

Example 12. *Consider the r-hybrid KB $\mathcal{B} = (\Phi, P)$, where $\Phi = \{\neg C(a)\}$ and $P = \{r \leftarrow C(x); s \leftarrow not\ r\}$. $\mathcal{B}$ does not entail s: any (infinite) SNA interpretation $\mathcal{I}$ in which r and $C(k)$ are true, for a single individual $k \neq a$, is an NM-model of $\mathcal{B}$: we have $\mathcal{I}|_{\Phi} \models \Phi$ and $\mathcal{I}|_P$ is a stable model of $\Pi(gr(P), \mathcal{I}|_{\Phi}) = \{r; s \leftarrow not\ r\}$. When considering the finite signature of $\mathcal{B}$, s belongs to every stable^E and every stable^A expansion of $\tau(P) \cup \Phi \cup CIA_{\Sigma_\Phi} \cup UNA_{\Sigma_P} \subseteq \mathcal{L}$, for $\tau \in \{\tau_{HP}, \tau_{EH}\}$.*

The counterexample does not apply when considering any- or all-name semantics with infinite signatures; resolving this and further studies will be done in future work.

Defining variants of MKNF with the any- or all-name semantics is straightforward. It remains to be seen whether the results on the embedding of MKNF into FO-AEL we presented in this paper extend to this case. Results on the comparison of embeddings of logic programs into the any-name semantics (de Bruijn *et al.* 2007) suggest that they do extend to hybrid MKNF KBs with safe rules.

Discussion and Conclusion

In this paper, we have considered possible embeddings of three major approaches to combining rules and ontologies, viz. dl-programs, r-hybrid KBs, and hybrid MKNF KBs, into first-order autoepistemic logic, which serves as a unifying framework in the sense that these approaches can be embedded into a single, well-studied formalism for knowledge representation and reasoning. Furthermore, by the embeddings and properties which they (perhaps necessarily) have, we also have a means for comparing various formalisms, and can assess their commonalities and differences.

In earlier work (de Bruijn *et al.* 2007), we have found that the behavior under combination with FO theories of the HP and EH embeddings are quite different. In the light of this it may be surprising that the embeddings of dl-programs and r-hybrid KBs are orthogonal to the choice between HP and EH: in dl-programs, the embedding of dl-atoms makes the regular literals independent from the FO theory; in the r-hybrid KB embedding, the CI axioms fix the interpretation of classical predicates in a stable expansion, similar to the projection operation in the r-hybrid KB semantics. In our previous work (de Bruijn *et al.* 2007), we considered also a third rule embedding, EB, which is like EH, but without the modal atom in the consequent and with PI axioms. We did not consider this embedding here for brevity. The results in this paper that apply to both HP and EH also extend to EB.

The difference between the embedding of dl-atoms and of the classical atoms in r-hybrid and hybrid MKNF KBs into FO-AEL suggests, not surprisingly, that there are fundamental differences in the interaction between the ontology and the rules that are not easily reconciled.

The fact that there is no faithful modular embedding of hybrid MKNF knowledge bases into FO-AEL (which fails vice versa) suggests that the former formalism is semantically more involved, and that further investigation is required to establish the difference between hybrid MKNF knowledge bases and combinations of rules and ontologies based on FO-AEL, as proposed by de Bruijn *et al.* (2007). In the same vein, the existence of modular embeddings of r-hybrid KBs into both FO-AEL and hybrid MKNF$^+$ KBs suggests that hybrid MKNF$^+$ KBs are semantically less involved than r-hybrid KBs.

In addition to the embeddings that we have presented, others may conceivably be considered, which are either defined from first principles or obtained by composing existing embeddings. For example, r-hybrid KBs may be embedded into FO-AEL by composing an embedding of r-hybrid into hybrid MKNF$^+$ KBs (see Motik & Rosati (2007b)) and the embedding of MKNF into FO-AEL.

Finding faithful modular embeddings for interesting fragments of (the strong answer set semantics for) dl-programs and hybrid MKNF$^+$ KBs into FO-AEL remains as an interesting issue, given the different underlying assumptions in the formalisms (groundedness in dl-programs vs. ungroundedness in AEL) and the negative results for instances with plain rule and ontology parts (for hybrids MKNF KBs).

Acknowledgements We thank the referees for their helpful comments on this paper. This work was partially supported by Austrian Science Fund (FWF) grant P20840.

References

Chen, J. 1993. Minimal knowledge + negation as failure = only knowing (sometimes). In *Proc. LPNMR'93*, 132–150.

Chen, J. 1994. The logic of only knowing as a unified framework for non-monotonic reasoning. *Fund. Inf.* 21:205–220.

de Bruijn, J.; Eiter, T.; Polleres, A.; and Tompits, H. 2006. On representational issues about combinations of classical theories with non-monotonic rules. In *Proc. KSEM'06*, LNCS 4092, 1-22.

de Bruijn, J.; Eiter, T.; Polleres, A.; and Tompits, H. 2007. Embedding non-ground logic programs into autoepistemic logic for knowledge-base combination. In *Proc. IJCAI'07*, 304–309.

de Bruijn, J.; Pearce, D.; Polleres, A.; and Valverde, A. 2007b. Quantified equilibrium logic and hybrid rules. In *Proc. RR'07*, LNCS 4524, 58–72. Springer.

Denecker, M.; Marek, V. W.; and Truszczynski, M. 2003. Uniform semantic treatment of default and autoepistemic logics. *Artificial Intelligence* 143(1):79–122.

Eiter, T.; Ianni, G.; Krennwallner, T.; and Schindlauer, R. 2007. Exploiting conjunctive queries in description logic programs. In *Proc. DL'07*, 259–266. Also ISAIM'08.

Eiter, T.; Lukasiewicz, T.; Schindlauer, R.; and Tompits, H. 2004. Combining answer set programming with description logics for the semantic web. In *Proc. KR'04*, 141–151.

Gelfond, M., and Lifschitz, V. 1991. Classical negation in logic programs and disjunctive databases. *New Generation Computing* 9(3-4):365–386.

Gottlob, G. 1995. Translating default logic into standard autoepistemic logic. *Journal of the ACM* 42(4):711–740.

Horrocks, I.; Patel-Schneider, P. F.; Bechhofer, S.; and Tsarkov, D. 2005. OWL rules: A proposal and prototype implementation. *Journal of Web Semantics* 3(1):23–40.

Janhunen, T. 1999. On the intertranslatability of non-monotonic logics. *Ann. Math. & Art. Int.* 27(1-4):79–128.

Konolige, K. 1991. Quantification in autoepistemic logic. *Fundamenta Informaticae* 15(3–4):275–300.

Levesque, H. J. 1990. All I know: A study in autoepistemic logic. *Artificial Intelligence* 42(2–3):263–309.

Lifschitz, V., and Schwarz, G. 1993. Extended logic programs as autoepistemic theories. In *Proc. LPNMR'93*, 101–114.

Lifschitz, V. 1991. Non-monotonic databases and epistemic queries. In *Proc. IJCAI 03*, 381–386. Morgan Kaufmann.

Marek, V. W., and Truszczyński, M. 1993. *Non-monotonic logic: Context-dependent reasoning*. Springer.

Moore, R. C. 1985. Semantical considerations on non-monotonic logic. *Artificial Intelligence* 25(1):75–94.

Motik, B., and Rosati, R. 2007a. A faithful integration of description logics with logic programming. In *Proc. IJCAI'07*, 477–482.

Motik, B., and Rosati, R. 2007b. Closing semantic web ontologies. Tech. Rep., Univ. Manchester.

Reiter, R. 1987. A logic for default reasoning. *Artificial Intelligence* 13:81–132.

Rosati, R. 1997. Embedding minimal knowledge into autoepistemic logic. In *Proc. AI*IA'97*, LNCS 1321, 231–241.

Rosati, R. 2005. On the decidability and complexity of integrating ontologies and rules. *Journal of Web Semantics* 3(1):61–73.

Rosati, R. 2006. $\mathcal{DL}+log$: Tight integration of description logics and disjunctive datalog. In *Proc. KR'06*, 68–78.

Schwarz, G. 1996. On embedding default logic into Moore's autoepistemic logic. *Artificial Intelligence* 80(1-2):349–359.