ELSEVIER

Artificial Intelligence 80 (1996) 349-359

Artificial Intelligence

Research Note

On embedding default logic into Moore's autoepistemic logic

Grigori Schwarz [*]

2747 Del Medio Ct., #108, Mountain View, CA 94040, USA

Received June 1995; revised November 1995

Abstract

Recently Gottlob proved [2] that there does not exist a faithful modular translation of default logic into autoepistemic logic, and presented a non-modular translation. Gottlob's translation, however, is indirect (it uses "nonmonotonic logic N" as an intermediate point), quite complex and exploits sophisticated encoding of proof theory in autoepistemic formulas. We provide a simpler and more intuitive (non-modular) direct translation. In addition, our argument is purely model-theoretic.

1. Introduction

Since Konolige's paper [4] much work has been done in attempts to provide a faithful translation of default logic [12] into Moore's autoepistemic logic [11]. The reason, apparently, was, that at the first look, Moore's logic seems to be clearly more general than Reiter's default logic: it, in a sense, "descends" defaults from the meta-level to the object level. In Reiter's logic, defaults are rules of inference; the language of Moore's logic is the usual language of classical modal logic, and the default rules are represented as usual modal formulas.

Konolige [4] suggested the following translation of default logic into the modal language. To each default of the form

$$d = \frac{\varphi : \psi_1, \ldots, \psi_m}{\eta} \tag{1}$$

he assigns a modal formula $L\varphi \wedge M\psi_1 \wedge \cdots \wedge M\psi_m \supset \eta$.

[*] E-mail: schwarz@flamingo.stanford.edu.

0004-3702/96/$15.00 © 1996 Elsevier Science B.V. All rights reserved
SSDI 0004-3702(95)00117-4

Unfortunately, it turned out that Konolige's translation, although being intuitive, does not give a faithful embedding of default logic into Moore's logic. In order to get the faithfulness, Konolige introduced the notion of a "strongly grounded expansion", which is not very intuitive. It is not invariant w.r.t. propositional equivalence, and, basically, mimics the rule-based nature of Reiter's logic, thus eliminating main advantages of Moore's logic—its uniformity and purely logical character.

Moore's logic was introduced as a refinement of nonmonotonic modal logics introduced by McDermott and Doyle [9] and later generalized by McDermott [8]. Later [16] it was shown that, formally, Moore's logic is a special case of McDermott-Doyle style nonmonotonic modal logics, namely it is nonmonotonic logic based on the modal logic KD45. The family of McDermott-Doyle style nonmonotonic logics was investigated in [6, 13, 16], and it turns out that by changing the underlying monotonic modal logic we get nonmonotonic logics which work not worse than Moore's autoepistemic logic, and overcome many of the difficulties of the original Moore's logic. Truszczyński [17] modified Konolige's translation and showed that the modified translation faithfully embeds default logic into nonmonotonic modal logic $\mathcal{S}$, for a wide class of modal logics $\mathcal{S}$, namely, for each $\mathcal{S}$ contained in the logic S4F and enjoying the necessitation rule. Logics of this kind include, in particular, such popular logics as S4, K, T.

Thus, in this respect (and in some others too—see [14, 16]) Moore's autoepistemic logic is among a few inappropriate nonmonotonic modal logics, whereas almost every other nonmonotonic modal logic would work.

Nevertheless, Moore's autoepistemic logic still remains the most popular among nonmonotonic modal logics. I think that the reason is, besides obvious historical reasons, that it is defined in terms of usual propositional consequence, without referring to any particular modal system.

Gottlob [2] demonstrated that the "purely propositional" nature of the autoepistemic logic is, probably, the main reason for difficulties in embedding the default logic into Moore's autoepistemic logic. Namely, a translation is called *modular* if for any default theory, the translation of the theory is the union of translations of all its defaults and axioms. Gottlob proved that a modular faithful translation of default logic into Moore's autoepistemic logic does not exist. Because researchers are usually looking for a modular translation, and modular translations are most natural, the reason for failure to find a translation of the default logic into Moore's autoepistemic logic becomes clear. And the crucial point in Gottlob's proof is the use of the deduction theorem for propositional logic. Because the deduction theorem in its usual form fails for modal logics, it is clear that the "propositional nature" of Moore's logic is essential here.

At the same time, Gottlob presented a non-modular translation of default logic into autoepistemic logic. In other words, we have to transform the default theory as a whole, not as a sum of translations of its elements.

Gottlob's translation is based on the following idea. Truszczyński's translation faithfully transforms a default theory $\mathcal{D}$ into a modal theory $tr(\mathcal{D})$. A theory T is an extension of $\mathcal{D}$ if and only if T is the objective part of an $\mathcal{S}$-expansion of $tr(\mathcal{D})$, for $\mathcal{S}$ being the modal logic N of "pure necessitation".

Gottlob builds a non-modular translation of nonmonotonic N into Moore's a.e. logic (nonmonotonic KD45), and takes the composition of the two translations. The translation of nonmonotonic N into nonmonotonic KD45 has the following form: For a given finite set of formulas Σ, the translation of Σ is a theory $\Sigma \cup \{LG(\Sigma)\}$, where $G(\Sigma)$ is a complicated formula, which includes additional propositional variables and encodes in some tricky way deductive properties of the modal logic N. The conjunct $LG(\Sigma)$, in a sense, "destroys" all ungrounded stable expansions of Σ, while preserving the "grounded" expansions (i.e. those which are also expansions in nonmonotonic modal logic N). The formula is quite complex, and for a person unfamiliar with the involved proof-theoretic machinery of modal logic N may be difficult to grasp.

From a mathematical point of view, the main contribution of Gottlob's [2] is a translation of nonmonotonic logic N into autoepistemic logic. Nonmonotonic N is a powerful formalism, which was studied by Marek and Truszczyński [7] and by Schwarz and Truszczyński [15]. Although there exists a model-theoretic characterization of nonmonotonic N [15], it is complicated, involves infinitely many accessibility relations, so the use of complicated proof-theoretic machinery by Gottlob seems to be justified.

The goal of the present note is to show that if we give up modularity, allow the use of new variables and do not wish to translate infinite theories, then we can come up with a translation of default logic into autoepistemic logic which is much simpler than that of Gottlob. Our translation is still quadratic in size and uses auxiliary variables, but it is much more intuitive and is, basically, a generalization of Konolige's translation (more exactly, of Chen's variant of the translation). Plus, our argument is purely model-theoretic.

2. Preliminaries

We assume familiarity with the basics of default logic [12] and autoepistemic logic [11]. A reader unfamiliar with these logics may treat results on semantical characterization of extensions (stable expansions) of default (autoepistemic) theories, which are reproduced below, as definitions.

Let us recall that a *default theory* is a pair (W, D), where W is a set of propositional formulas, and D is a set of syntactical constructs of the form (1) (called *default rules* or *defaults*). Formula φ is called *prerequisite* of the default (1), ψ_i's are called *justifications* and η is called the *conclusion* of (1). A default can contain no prerequisite, in which case φ is absent from (1), or no justifications, which correspond to $m = 0$ in (1). In this paper we, following Gottlob [2], always assume W and D to be finite.

A language of autoepistemic logic is obtained from the propositional language by adding a modal belief operator L. A modal operator M is an abbreviation for $\neg L \neg$.

We assume that our basic propositional language contains infinitely many propositional variables. So when we speak about "new propositional variables" in the context of a given set of formulas, we mean variables from our language which do not appear in the formulas under discussion in the context.

First, we recall a semantical characterization of default logic by Guereiro and Casanova [3].

A (propositional) interpretation (or a *world*) is an assignment of usual truth values to all propositional variables. Let (I, J) be a pair of sets of propositional interpretations. A default $d = \frac{\varphi:\psi_1,\ldots,\psi_m}{\eta}$ is *true* in (I,J) if, whenever φ is true in all the interpretations of J, and each of ψ_i is true in some interpretation in I, then η is true in all the interpretations in J. A formula η is true in (I,J) if it is true in all the interpretations in J. We will use the notation $(I,J) \models d$ to denote that a default (or a formula) d is true in (I,J), and $(I,J) \models \mathcal{D}$ to denote that all defaults and axioms of $\mathcal{D}$ are true in (I,J). By $I \models \varphi$ we denote that a propositional formula φ is valid in all the interpretations from I.

A set of interpretations I is a *model of a default theory* $\mathcal{D}$ if I is a maximal set J such that $(I,J) \models \mathcal{D}$.

Proposition 2.1 ([3]). *T is an extension of $\mathcal{D}$ if and only if $T = \{\varphi : I \models \varphi\}$ for some model I of $\mathcal{D}$.*

Notice that Proposition 2.1 includes the case of an inconsistent extension, when $I = \emptyset$.

Clearly, I is a default model of $\mathcal{D}$ if and only if $(I,I) \models \mathcal{D}$, and for no $J \supset I$, $(I,J) \models \mathcal{D}$. The following simple lemma shows that we can restrict the cardinalities of the J's.

Lemma 2.2. *Let $\mathcal{D}$ be a default theory containing k defaults with prerequisites. Then I is a default model of $\mathcal{D}$ if and only if $(I,I) \models \mathcal{D}$, and for all $J \supset I$ such that the cardinality of $J \setminus I$ does not exceed* $\max(1,k)$, $(I,J) \not\models \mathcal{D}$.

Proof. It is sufficient to show that for each $K \supset I$ such that $(I,K) \models \mathcal{D}$, there is J such that $I \subset J \subseteq K$, $|J \setminus I| \leqslant \max(1,k)$ and $(I,J) \models \mathcal{D}$. For each prerequisite φ of a default in D such that for some $\alpha \in K \setminus I$, $\alpha \not\models \varphi$, fix some α_φ with this property. Let J be the union of I and the set of all such α_φ; if for all prerequisites of all defaults, and for all $\alpha \in K \setminus I$, $\alpha \models \varphi$, then put $J = I \cup \{\alpha\}$ for an arbitrary α in $K \setminus I$. It is straightforward to show that $(I,J) \models \mathcal{D}$. □

A similar characterization exists for autoepistemic logic, tracing back to Moore [10], and appearing in a more clear form in Levesque [5]. Namely, for a modal formula φ, an interpretation α and a set of interpretations I (not necessarily containing α), the relation $(I,\alpha) \models \varphi$ is defined in the following natural way: For an atom p, $(I,\alpha) \models p$ iff p is true in α; $(I,\alpha) \models \varphi \wedge (\vee) \psi$ if $(I,\alpha) \models \varphi$ and (or) $(I,\alpha) \models \psi$; $(I,\alpha) \models \neg\psi$ if $(I,\alpha) \not\models \psi$; finally, $(I,\alpha) \models L\varphi$ iff for each $\beta \in I$, $(I,\beta) \models \varphi$. As usual, for an autoepistemic formula φ and a set I of propositional interpretations, $I \models L\varphi$ means that for each $\alpha \in I$, $(I,\alpha) \models L\varphi$. A set of interpretations, I, is called an *autoepistemic model of an autoepistemic theory* A, if $I = \{\alpha : (I,\alpha) \models A\}$.

Proposition 2.3 ([5,10]). *T is a stable expansion of A if and only if for some autoepistemic model I of A, $T = \{\varphi : I \models \varphi\}$.*

Again, the case of an inconsistent stable expansion is captured here by $I = \emptyset$.

Next we recall Chen's translation [1]. This is the following modification of the Konolige translation: Default rule (1) is translated into the autoepistemic formula

$$L\varphi \wedge \varphi \wedge M\psi_1 \wedge \cdots \wedge M\psi_m \supset \eta. \tag{2}$$

Chen proved that this translation is faithful for prerequisite-free default theories (which is true for Konolige's translation too—for prerequisite-free theories the two translations coincide), and for default theories in which all the prerequisites and conclusions are conjunctions of literals. In fact, by using the technique of this paper, it is possible to show that Chen's translation is faithful if all the prerequisites are conjunctions of literals, or if a default theory contains no more than one default with a prerequisite.

We will denote the formula (2) by $C(d)$, if d is a default rule (1). (If a default has no prerequisite, then its Chen's translation coincides with Konolige's translation.) Given the default theory (W, D) Chen's translation $C(D)$ is defined as $W \cup \{C(d) : d \in D\}$. If both W and D are finite, then we denote by $C(\mathcal{D})$ the conjunction of W with all the members of $C(D)$.

The following simple lemma immediately follows from the definitions and explains why we have to use Chen's translation rather than Konolige's one.

Lemma 2.4. *Let I be a set of interpretations, let $\mathcal{D}$ be a default theory, α be an arbitrary propositional interpretation. Then $(I, I \cup \{\alpha\}) \models \mathcal{D}$ if and only if $(I, \beta) \models C(\mathcal{D})$ for each $\beta \in I \cup \{\alpha\}$.*

3. New translation

Let $\mathcal{D} = (W, D)$ be a finite default theory. Because W is finite, we can assume that W is just one formula. Let D contain exactly $k+1$ defaults with prerequisites, $k \geqslant -1$. That is D can also have any number of prerequisite-free defaults.

Let $p_1, \ldots, p_n$ be a complete list of propositional variables appearing in $\mathcal{D}$. Let us fix $\max(k, 0)$ additional lists of new propositional variables, $p_1^i, \ldots, p_n^i$, $1 \leqslant i \leqslant k$. (That is, if all the defaults except maybe one, are prerequisite-free, we do not have any additional variables; otherwise for each default with a prerequisite in excess of one, we create copies of original variables.)

For a formula ζ, all of whose variables are among $p_1, \ldots, p_n$, by $\zeta(p^i)$ we denote the result of the substitution of the variables $p_1^i, \ldots, p_n^i$ for the variables $p_1, \ldots, p_n$ in ζ. Sometimes ζ will be denoted by $\zeta(p)$ for better clarity. For each default $d \in D$,

$$d = \frac{\varphi : M\psi_1, \ldots, M\psi_m}{\eta},$$

by $t(d)$ we denote

$$(L\varphi \wedge \varphi \wedge \varphi(p^1) \wedge \cdots \wedge \varphi(p^k) \wedge M\psi_1 \wedge \cdots \wedge M\psi_m) \supset$$
$$(\eta \wedge \eta(p^1) \wedge \cdots \wedge \eta(p^k)).$$

(If the default is prerequisite-free, we do not have premises $L\varphi \wedge \varphi \wedge \cdots \wedge \varphi(p^k)$ in this formula.)

We are ready to define our translation $t(\mathcal{D})$:

$$t(\mathcal{D}) = C(\mathcal{D}) \vee \left(W \wedge W(p^1) \wedge \cdots \wedge W(p^k) \wedge \bigwedge_{d \in D} t(d) \right).$$

Example 3.1. Let a default theory consist of the single default $\frac{p:}{p}$—a famous example showing that Konolige's translation does not work. Our translation of this theory is $(Lp \wedge p \supset p) \vee (Lp \wedge p \supset p)$ (clearly, if we have a single default, our translation does not have any additional variables). Compare with Gottlob's translation of this theory:

$$\begin{aligned}
&(Lp \supset p) \wedge \\
&L((((([Lp^z] \supset p^z) \wedge (\neg Lp \supset \neg [Lp^z]) \wedge (u_p \supset [Lp^z])) \supset (Lp \supset [Lp^z])).\vee. \\
&\quad (\neg u_p \wedge ((([Lp^p] \supset p^p) \wedge (\neg Lp \supset \neg [Lp^p]) \wedge (u_p \supset [Lp^p])) \supset p^p))),
\end{aligned}$$

where $[Lp^z]$, p^z, u_p, $[Lp^p]$ and p^p are new propositional variables.

Example 3.2. Let $\mathcal{D}_2$ consist of two defaults: $\frac{p \supset q:}{p}$ and $\frac{p:}{q}$. For this theory, neither Konolige's nor Chen's translation is faithful. Chen's translation is the conjunction of $(L(p \supset q) \wedge (p \supset q) \supset p)$ and $Lp \wedge p \supset q$, and has two stable expansions, one of them containing p and q is clearly "ungrounded", and the default theory in question has only one extension (set of all tautologies). Our translation of this default theory is the disjunction of Chen's translation and the formula

$$(L(p \supset q) \wedge (p \supset q) \wedge (p_1 \supset q_1) \supset p \wedge p_1) \wedge (Lp \wedge p \wedge p_1 \supset q \wedge q_1).$$

This theory does not have an expansion containing p and q—additional variables in the second disjunct "block" the inference of $p \supset q$ and p from Lp and $L(p \supset q)$ (which was possible for Chen's translation). Note that Gottlob's translation of this theory is very long and contains 11 additional variables versus just the two we have.

The above examples do not illustrate the role of the disjunctive term $C(\mathcal{D})$ in the translation. Let us consider the default theory of the last example augmented by the axiom p. It has the unique extension $Th(p,q)$. Chen's translation of this theory is obtained from Chen's translation of the theory of the last example by just adding the conjunct p. The second disjunct of our translation is like in the previous example, plus the conjunct $p \wedge p_1$. If we had not added disjunctively Chen's translation, we would still have only one stable expansion, but it would contain p_1 and q_1 besides p and q, respectively. Adding the Chen disjunct "blocks" the inference of p_1 and q_1, while still allowing to derive p and q, which are implied by both the disjuncts of our translation.

We see that our proposed translation bears a clear structural relationship to the source theory, and is structurally similar to Konolige's and Chen's translations which both have a good underlying intuition.

Theorem 3.3. *Let $\mathcal{D}$ be a default theory. Then for each set of propositional sentences T, T is an extension of $\mathcal{D}$ if and only if T is the objective part of some stable expansion of $t(\mathcal{D})$.*

For the proof of Theorem 3.3, we need first to prove some lemmas. The following lemma immediately follows from our definitions and illustrates the role of Chen's disjunct in our translation.

Lemma 3.4. *Let $\mathcal{D} = (W, D)$ be a finite default theory, let $t(\mathcal{D})$ be the translation of $\mathcal{D}$ as defined above. Let I be an autoepistemic model of $t(\mathcal{D})$. Then for all $\alpha \in I$, $(I, \alpha) \models C(\mathcal{D})$.*

Proof. Let $p_1, \ldots, p_n$ be a complete list of propositional variables occurring in $\mathcal{D}$. Assume, to the contrary, that for some $\alpha \in I$, $(I, \alpha) \not\models C(\mathcal{D})$. Then the second disjunct of $t(\mathcal{D})$ must hold, that is $(I, \alpha) \models W \wedge W(p^1) \wedge \cdots$—hence, $\alpha \models W$, and we must have that for some default $d \in D$, $(I, \alpha) \not\models C(d)$. Assume that d is of the form (1). Because $(I, \alpha) \models t(\mathcal{D})$, we necessarily have

$$(I, \alpha) \models L\varphi \wedge \varphi \wedge \varphi(p^1) \wedge \cdots \wedge \varphi(p^k) \wedge M\psi^1 \wedge \cdots \wedge M\psi^m \supset \eta \wedge \eta(p^1) \wedge \cdots \wedge \eta(p^k) \tag{3}$$

and at the same time,

$$(I, \alpha) \models L\varphi \wedge \varphi \wedge M\psi^1 \wedge \cdots \wedge M\psi^m \tag{4}$$

and $(I, \alpha) \not\models \eta$.

Hence the only way (3) can be true is, if for some $j \leqslant k$, $(I, \alpha) \not\models \varphi(p^j)$. Let us construct a propositional valuation β as follows. Let $\beta(p_l) = \alpha(p_l^j)$ for each $1 \leqslant l \leqslant n$, and let $\beta(p) = \alpha(p)$ for all other variables p. We thus have $\beta \not\models \varphi$. Hence, and from (4) (because $(I, \alpha) \models L\varphi$), we conclude that $\beta \notin I$.

Furthermore, for each $l \leqslant k$ we have $\beta \models W(p^l)$, and $\beta \models W$ (because $\alpha \models W(p^j)$). We also can prove that for each default $d \in D$, $(I, \beta) \models t(d)$. Indeed, let $t(d)$ have the form

$$L\rho \wedge \rho(p) \wedge \rho(p^1) \wedge \cdots \wedge \rho(p^k) \wedge M\tau_1 \wedge \cdots \wedge M\tau_m \supset \zeta(p) \wedge \zeta(p^1) \wedge \cdots \wedge \zeta(p^k).$$

Since I is an autoepistemic model of $t(\mathcal{D})$, we have $(I, \alpha) \models t(d)$. If this relation holds because the consequent of $t(d)$ is true in α, then it is also true in β, because $\beta(\zeta(p)) = \alpha(\zeta(p^j))$. On the other hand, if this relation holds because one of $M\tau_s$ is false in (I, α), then it is trivially false in (I, β), because the truth value of $M\tau_s$ depends only on I. Further, if this relation holds because for some l, $\alpha \not\models \rho(p^l)$, then of course $\beta \not\models \rho(p^l)$, as α agrees with β on these variables. Finally, if $\alpha \not\models \rho$ or $(I, \alpha) \not\models L\rho$, then also $(I, \alpha) \not\models L\rho$ (because $\alpha \in I$), so we have $(I, \beta) \not\models L\rho$, and $(I, \beta) \models t(d)$.

Thus, we have

$$(I, \beta) \models W \wedge W(p^1) \wedge \cdots \wedge W(p^k) \wedge \bigwedge_{d \in D} t(d),$$

hence $(I, \beta) \models t(\mathcal{D})$, but I is an autoepistemic model of $t(\mathcal{D})$, so we must have $\beta \in I$—a contradiction. □

Lemma 3.5. *Let $\mathcal{D} = (W, D)$ be a finite default theory all whose propositional variables are among $p_1, \ldots, p_n$. Let $t(\mathcal{D})$ be the translation of $\mathcal{D}$ as defined above. Let I be an autoepistemic model of $t(\mathcal{D})$, and let $\alpha \in I$. Let β be a propositional interpretation which agrees with α on all propositional letters from the list $P = p_1, \ldots, p_n$. Then $\beta \in I$.*

Proof. Assume $(I, \alpha) \models t(\mathcal{D})$. Then, by Lemma 3.4, $(I, \alpha) \models C(\mathcal{D})$. Because $C(\mathcal{D})$ contains only variables from $\{p_1, \ldots, p_n\}$, we have $(I, \beta) \models C(\mathcal{D})$, hence $(I, \beta) \models t(\mathcal{D})$, hence $\beta \in I$. □

The following is the semantical analogue of the main theorem.

Lemma 3.6. *I is a model of a default theory $\mathcal{D}$ if and only if I is an autoepistemic model of $t(\mathcal{D})$.*

Proof. Let us prove the "only if" part first.

Assume that I is a model of $\mathcal{D}$. By Lemma 2.4, for all $\alpha \in I$, $(I, \alpha) \models C(\mathcal{D})$, and $C(\mathcal{D})$ is just the first disjunct of $t(\mathcal{D})$, so $(I, \alpha) \models t(\mathcal{D})$.

Now, assume that for some $\alpha \notin I$, $(I, \alpha) \models t(\mathcal{D})$. Then $(I, \alpha) \models C(\mathcal{D})$ or

$$(I, \alpha) \models W \wedge W(p^1) \wedge \cdots \wedge W(p^k) \wedge \bigwedge_{d \in D} t(d).$$

Case 1: $(I, \alpha) \models C(\mathcal{D})$. Then consider $J = I \cup \{\alpha\}$. By Lemma 2.4, $(I, J) \models \mathcal{D}$, which contradicts the assumption that I is a default model of $\mathcal{D}$.

Hence, we must have:

Case 2:

$$(I, \alpha) \models W \wedge W(p^1) \wedge \cdots \wedge W(p^k) \wedge \bigwedge_{d \in D} t(d). \tag{5}$$

Let us construct a set of interpretations J as follows. Let α_0 coincide with α, and for $1 \leqslant j \leqslant k$ ($k + 1$ is the number of defaults with prerequisites), let $\alpha_j(p_s) = \alpha(p_s^j)$, $1 \leqslant s \leqslant n$ (n is the number of variables occurring in $\mathcal{D}$), and for the remaining variables α_j may be arbitrary. Put $J = I \cup \{\alpha_0, \ldots, \alpha_k\}$. Because $\alpha_0 \notin I$, J is a proper superset of I. We claim that $(I, J) \models \mathcal{D}$.

First, we have to show that $(I, J) \models W$. For all $\gamma \in I$, $\gamma \models W$ by our assumption. Also, α_0 coincides with α on all the variables in W, so $\alpha_0 \models W$ by (5). By definition of α_i for $i \geqslant 1$, $\alpha_i(W(p)) = \alpha(W(p^i))$, so from (5) we have $\alpha_i \models W$.

Now assume $d \in D$, d of the form (1). Let us prove that $(I, J) \models d$. Assume for all $\beta \in J$, $\beta \models \varphi$. First of all, it follows immediately that $(I, \alpha) \models L\varphi$ holds. Then we have for each α_i, $\alpha_i \models \varphi$, hence $\alpha \models \varphi(p^i)$. Also, because $\alpha_0 \models \varphi$, we have $\alpha \models \varphi$. If, in addition, for each ψ_j there is a world in I making ψ_j true, we also have $(I, \alpha) \models M\psi_j$. Since we also have $(I, \alpha) \models t(d)$, it follows that the consequent of $t(d)$ is true in α, so $\alpha \models \eta \wedge \eta(p^1) \wedge \cdots \wedge \eta(p^k)$, hence for each j we have $\alpha_j \models \eta$. Finally, since I is a model of $\mathcal{D}$, we also have $\gamma \models \eta$ for all $\gamma \in I$, and this concludes the proof of $(I, J) \models d$. Because d was an arbitrary default in D, we have $(I, J) \models \mathcal{D}$—a contradiction with the assumption that I is a default model of $\mathcal{D}$, which concludes the proof of the "only if" part.

For the "if" part, assume that I is an autoepistemic model of $t(\mathcal{D})$. From Lemma 3.4 it follows that $(I,\alpha) \models C(\mathcal{D})$ for all $\alpha \in I$, hence by Lemma 2.4, $(I,I) \models \mathcal{D}$. Now, assume that I is not a default model of $\mathcal{D}$. Then by Lemma 2.2, for some set F of propositional interpretations disjoint with I, $F = \{\alpha_0,\dots,\alpha_l\}$, $0 \leqslant l \leqslant \max(0,k)$, $(I, I \cup F) \models \mathcal{D}$. Then we construct an interpretation α as follows. Let $\alpha(p_i) = \alpha_0(p_i)$; for $j \leqslant l$, let $\alpha(p_i^j) = \alpha_j(p_i)$; and finally, for $l < j \leqslant k$, let $\alpha(p_j^i) = \alpha_0(p_i)$, where $1 \leqslant l \leqslant n$.

For each p_i, $\alpha(p_i) = \alpha_0(p_i)$, hence, by Lemma 3.5, $\alpha \notin I$. Let us show that $(I,\alpha) \models t(\mathcal{D})$, which will complete the proof by contradiction. We will show that the second disjunct in $t(\mathcal{D})$ holds in (I,α). First by definition of α, we have $\alpha(p_i^j) = \alpha_j(p_i)$ for $0 < j \leqslant l, 0 \leqslant i \leqslant n$, so we have $\alpha \models W(p^j)$. Similarly, $\alpha \models W(p)$ and $\alpha \models W(p^j)$ for all j, $l < j \leqslant k$. So it remains to prove that for all the defaults $d \in D$, $(I,\alpha) \models t(d)$. But if all the premises of $t(d)$ are true, then we easily conclude that φ is true in all the worlds in $I \cup F$, and each of ψ_i is true somewhere in I, and, because $(I, I \cup F) \models d$, we obtain that for each i, $\alpha_i \models \eta$, which implies, by our construction of α, that $\alpha \models \eta \wedge \eta(p^1) \wedge \cdots \wedge \eta(p^k)$, so $(I,\alpha) \models t(\mathcal{D})$—a contradiction with the assumption that I is an autoepistemic model of $t(\mathcal{D})$. □

Theorem 3.3 now immediately follows from Propositions 2.1 and 2.3 and Lemma 3.6.

The idea of our construction is very simple. For models of default logic we have to consider an additional set of interpretations (worlds), whereas for autoepistemic models only one additional world is allowed. We create copies of all the variables in $\mathcal{D}$, one copy for each additional world, and simulate the whole cluster within this one additional world.

Let us call a translation t *locally faithful* if for each $\mathcal{D}$ containing only variables from a given list P, T is a stable expansion of $t(\mathcal{D})$ if and only if the restriction of T to the propositional language based on P is an extension of $\mathcal{D}$. Then we can drop the disjunct $C(\mathcal{D})$ from our translation, and the simplified translation remains locally faithful. (The proofs are essentially the same, even simpler, as we do not need counterparts of Lemmas 3.4 and 3.5.) This way, the translation becomes "locally modular" in a sense that it is modular for theories with the number of default rules bounded by a given constant.

4. Conclusion

We constructed a non-modular translation of default logic into autoepistemic logic which, we believe, is easier to grasp than Gottlob's translation. In addition, our motivations and proofs are purely model-theoretic, whereas Gottlob's approach was purely proof-theoretic. Like Gottlob's translation, our translation introduces new variables, and is quadratic in size, but the multiplicative constant is much lower. Of course, our translation does not make Gottlob's translation obsolete; the main achievement of Gottlob was to translate nonmonotonic logic N (which is, in a sense, the "most grounded" nonmonotonic modal logic) into autoepistemic logic. Our aim was to show that if our goal is more modest, then it can be achieved with simpler means.

It is, apparently, still an open problem whether there exists a non-modular polynomial translation of default logic into autoepistemic logic which does not introduce new variables.

Acknowledgment

The author is grateful to Georg Gottlob, who carefully read earlier versions of this paper and made numerous useful comments. In particular, initially all the results were formulated for consistent extensions and consistent stable expansions; Georg noticed that the results remain true (and the proofs continue to hold) in the general case.

The author is grateful to anonymous referees for exceptionally thorough reviewing and numerous suggestions on improving the paper. (The report of one of the referees was 15 pages long!)

The work was performed when the author was visiting the Computer Science Department of the University of California, Riverside.

References

[1] J. Chen, The logic of only knowing as a unified framework of nonmonotonic reasoning, *Fund. Inform.* **21** (1994) 205-220.

[2] G. Gottlob, The power of beliefs or translating default logic into standard autoepistemic logic, in: *Proceedings IJCAI-93*, Chambery, France (1993) 570-575; extended version: *J. ACM* **42** (4) (1993) 711-740.

[3] R. Guerreiro and M. Casanova, An alternative semantics for default logic, Preprint, The Third International Workshop on Nonmonotonic Reasoning, South Lake Tahoe (1990).

[4] K. Konolige, On the relation between default and autoepistemic logic, *Artif. Intell.* **35** (1988) 343-382.

[5] H.J. Levesque, All I know: a study in autoepistemic logic, *Artif. Intell.* **42** (1990) 263-309.

[6] W. Marek, G. Schwarz and M. Truszczyński, Modal nonmonotonic logics: ranges, characterization, computation, *J. ACM* **40** (1993) 963-990.

[7] W. Marek and M. Truszczyński, Modal logic for default reasoning, *Ann. Math. Artif. Intell.* **1** (1990) 275-302.

[8] D. McDermott, Nonmonotonic logic II: nonmonotonic modal theories, *J. ACM* **29** (1982) 33-57.

[9] D. McDermott and J. Doyle, Nonmonotonic logic I, *Artif. Intell.* **13** (1980) 41-72.

[10] R.C. Moore, Possible-world semantics autoepistemic logic, in: R. Reiter, ed., *Proceedings of the Workshop on Non-Monotonic Reasoning* (1984) 344-354; reprinted in: M. Ginsberg, ed., *Readings on Nonmonotonic Reasoning* (Morgan Kaufmann, San Mateo, CA, 1987) 137-142.

[11] R.C. Moore, Semantical considerations on non-monotonic logic, *Artif. Intell.* **25** (1985) 75-94.

[12] R. Reiter, A logic for default reasoning, *Artif. Intell.* **13** (1980) 81-132.

[13] G. Schwarz, Bounding introspection in nonmonotonic logic, in: C. Rich, B. Nebel and W. Swartout, eds., *Principles of Knowledge Representation and Reasoning, Proceedings of the 3rd International Conference* (*KR-92*) (Morgan Kaufmann, San Mateo, CA, 1992) 581-590.

[14] G. Schwarz and M. Truszczyński, Modal logic S4F and the minimal knowledge paradigm, in: Y. Moses, ed., *Theoretical Aspects of Reasoning about Knowledge, Proceedings of the 4th Conference* (*TARK 1992*) (Morgan Kaufmann, San Mateo, CA, 1992) 184-198.

[15] G. Schwarz and M. Truszczyński, Subnormal logics for knowledge representation, in: *Proceedings AAAI-93*, Washington, DC (AAAI Press, 1993) 438-443.

[16] G.F. Shvarts (G. Schwarz), Autoepistemic modal logics, in: R. Parikh, ed., *Theoretical Aspects of Reasoning about Knowledge, Proceedings of the 3rd International Conference* (*TARK 1990*) (Morgan Kaufmann, San Mateo, CA, 1990) 97–109.

[17] M. Truszczyński. Embedding default logic into modal nonmonotonic logics. in: W. Marek, A. Nerode and V.S. Subramahmanian, eds., *Logic Programming and Non-Monotonic Reasoning. Proceedings of the First International Workshop* (MIT Press, Cambridge, MA, 1991) 151–165.