ELSEVIER

Artificial Intelligence 77 (1995) 285–319

Artificial Intelligence

A comparative study of open default theories

Michael Kaminski*

Department of Computer Science, The Hong Kong University of Science and Technology, Clear Water Bay Road, Kowloon, Hong Kong

Received February 1993; revised December 1993

Abstract

The paper examines the definitions of open default theories known from the literature. First it is shown that none of them is satisfactory either for formal or for intuitive reasons. Next a new approach is considered. It is free from the obvious deficiencies of the known definitions, but possesses their positive properties.

1. Introduction

One of the widely used nonmonotonic formalisms is Reiter's *default logic* [15]. This logic deals with rules of inference called *defaults* which are expressions of the form

$$\delta(\boldsymbol{x}) = \frac{\alpha(\boldsymbol{x}) : M\beta_1(\boldsymbol{x}), \ldots, M\beta_m(\boldsymbol{x})}{\gamma(\boldsymbol{x})},$$

where $\alpha(\boldsymbol{x})$, $\beta_1(\boldsymbol{x}), \ldots, \beta_m(\boldsymbol{x})$, $\gamma(\boldsymbol{x})$, $m \geqslant 1$, are formulas of the first-order predicate calculus whose free variables are among $\boldsymbol{x} = x_1, \ldots, x_n$. A default is *closed* if none of $\alpha, \beta_1, \ldots, \beta_m, \gamma$ contains a free variable. Otherwise a default is called *open*. The formula $\alpha(\boldsymbol{x})$ is called the *prerequisite* of the default rule, the formulas $\beta_1(\boldsymbol{x}), \ldots, \beta_m(\boldsymbol{x})$ are called the *justifications*, and the formula $\gamma(\boldsymbol{x})$ is called the *conclusion*. Roughly speaking, the intuitive meaning of an open default is as follows. For every n-tuple of objects $\boldsymbol{t} = t_1, \ldots, t_n$, if $\alpha(\boldsymbol{t})$ is believed, and the $\beta_i(\boldsymbol{t})$'s are consistent, then one is permitted to deduce $\gamma(\boldsymbol{t})$. Thus an open default can be thought of as a kind of a "default scheme", where the free variables $\boldsymbol{x}$ can

* Current address: Department of Computer Science, Technion–Israel Institute of Technology, Haifa 32000, Israel.

SSDI 0004-3702(94)00035-Y

be replaced by any of the theory's objects. Various examples of deduction by default rules can be found in [15].

Whereas closed defaults have been quite thoroughly investigated, very little is known about open ones. Moreover, there is no common attitude towards their meaning. However, interesting cases of default reasoning usually deal with open defaults, because the intended use of a default is to determine whether an object possesses a given property rather than accepting or rejecting a "fixed statement".

Three major approaches to the treatment of open defaults are known from the literature. The first one belongs to Reiter [15], where he gives explicit names to the theory objects by extending the theory language with new constants. Then Reiter treats an open default as a set of all its closed instances in the enriched language. The second approach is similar to the first one and belongs to Poole [14] who replaces an open default by the set of all its closed instances over the original language. The last approach is that of Lifschitz [9], where free variables in defaults are treated as object variables, rather than metavariables for the closed terms of the theory.

In this paper we examine the above approaches from formal and intuitive points of view. Obviously, the most natural formal test for accepting a definition of an extension for an open default theory is that it must be equivalent to the original definition of Reiter, when applied to a closed default theory. However, this necessary condition is not sufficient, because equivalent definitions of extensions for closed default theories become different when one extends them to open default theories. Since there are no (and cannot be any) formal criteria for a sufficient condition, in order to choose the right definition we should rely on our imprecise intuition to tell us what we should expect from an extension for an open default theory.

As the result of our analysis we argue that all Reiter's, Poole's and Lifschitz's definitions are not entirely sound. In particular, Reiter's definition that gives explicit names to implicitly defined objects is counterintuitive and also is not acceptable from a formal point of view, Poole's definition that deals only with explicitly defined objects is too weak (yet it passes the formal test), and Lifschitz's definition that treats explicitly defined objects as implicitly defined ones is not acceptable either for a formal or for an intuitive reason. However, Lifschitz's definition seems to be more promising, and we propose a modification to Lifschitz's definition which makes it free from its obvious deficiencies. The main feature of the modified definition is that it clearly separates between explicitly and implicitly defined theory objects. It also passes the formal test and its connection to circumscription is similar to the original one.

The paper is organized as follows. In the next section we recall the definition of extensions for closed default theories, in Sections 3 and 4 we examine, respectively, Poole's and Reiter's definitions of extensions for open default theories, and in Section 5 we consider a possible modification to Reiter's definition. Section 6 contains a semantical definition of extensions for closed default theories, which is the starting point for Lifschitz's and our approaches. In Section 7 we examine Lifschitz's approach to open default theories, in Section 8 we present a modification to this approach which is more robust than the original, and in Section 9 we show how extensions for default logic with fixed constants can be expressed in terms of the modified approach. In Section 10 we establish a relationship between the modified Lifschitz's approach and circumscription. Finally, we end the paper with some concluding remarks.

2. Closed default theories

In this section we recall Reiter's definition of extensions for closed default theories. This definition is frequently used in this paper as a formal criterion for accepting or rejecting definitions of extensions for open default theories. In particular, as was mentioned in the introduction, if we accept Reiter's definition of extensions for closed default theories as a "right one", then a "right" definition of extensions for open default theories, when applied to a closed default theory, must be equivalent to Reiter's definition.

Definition 1. A *default theory* is a pair (D, A), where D is a set of defaults and A is a set of first-order sentences (axioms). A default theory is called *closed*, if all its defaults are closed. Otherwise it is called *open*.

Definition 2. Let (D, A) be a closed default theory. For any set of sentences S let $\Gamma_{(D,A)}(S) = B$, where B is the smallest set of sentences (beliefs) that satisfies the following three properties.[1]

(D1) $A \subseteq B$.

(D2) $\mathrm{Th}(B) = B$, i.e., B is deductively closed.

(D3) If $\frac{\alpha\,:\, M\beta_1, \ldots, M\beta_m}{\gamma} \in D$, $\alpha \in B$, and $\neg\beta_1, \ldots, \neg\beta_m \notin S$, then $\gamma \in B$.

A set of sentences E is an *extension* for (D, A) if $\Gamma_{(D,A)}(E) = E$, i.e., if E is a fixed point of the operator $\Gamma_{(D,A)}$.

All the examples, but one, we consider in this paper deal with the simplest and very intuitive case of defaults of the form $\frac{:\, M\beta(x)}{\beta(x)}$ which are called *normal defaults without prerequisites*. We shall need the following lemma.

Lemma 3. *Let D be a set of closed normal defaults without prerequisites. Then a set of sentences is a consistent extension for (D, A) if and only if it is a maximal consistent set of sentences of the form* $\mathrm{Th}(A \cup A')$, *where* $A' \subseteq \{\beta\colon \frac{:\, M\beta}{\beta} \in D\}$.

Proof. The "if" part of the lemma is, actually, [14, Theorem 4.1] restricted to closed defaults.

[1] This definition follows [9] and differs from the original one in [15] in introducing the notation $\Gamma_{(D,A)}(S) = B$. This notation is more convenient for a technical reason.

For the "only if" part, let E be an extension for (D, A), and let $A_E = E \cap \{\beta: \frac{:M\beta}{\beta} \in D\}$. Then, by [15, Theorem 2.5], $E = \mathrm{Th}(A \cup A_E)$.

It remains to show that $\mathrm{Th}(A \cup A_E)$ $(=E)$ is a maximal consistent set of sentences of the form $\mathrm{Th}(A \cup A')$, where $A' \subseteq \{\beta: \frac{:M\beta}{\beta} \in D\}$. Assume to the contrary that this is not the case, and let $\mathrm{Th}(A \cup A')$, where $A' \subseteq \{\beta: \frac{:M\beta}{\beta} \in D\}$, be a maximal consistent set of sentences containing E as a proper subset. By the "if" part of the lemma, $\mathrm{Th}(A \cup A')$ is an extension for (D, A), in contradiction with the *Minimality of Extensions* [15, Theorem 2.4]. □

3. Poole's definition of open default theories

In this section we analyze Poole's definition of open default theories. Even though, this definition was introduced eight years later than Reiter's one, it is discussed first for a methodological reason. For both Poole's and Reiter's approaches we need the following definition.

Definition 4. A (closed) *instance* of an open default $\delta(\boldsymbol{x}) = \frac{\alpha(\boldsymbol{x}) : M\beta_1(\boldsymbol{x}), \ldots, M\beta_m(\boldsymbol{x})}{\gamma(\boldsymbol{x})}$ is a closed default $\delta(\boldsymbol{t}) = \frac{\alpha(\boldsymbol{t}) : M\beta_1(\boldsymbol{t}), \ldots, M\beta_m(\boldsymbol{t})}{\gamma(\boldsymbol{t})}$, where $\boldsymbol{t} = t_1, \ldots, t_n$ is a tuple of closed (or ground) terms of the underlying language $\mathbf{L}$. For an open default δ, the set of all closed instances of δ is denoted by $\bar{\delta}$, and for a set of defaults D, $\bar{D} = \bigcup_{\delta \in D} \bar{\delta}$ is the set of all closed instances of all defaults of D.

In [14] Poole deals with normal defaults without prerequisites only, and treats an open default theory (D, A) as a closed default theory $(\bar{D}, A)$.

Note that if all defaults from D are closed, then $\bar{D} = D$. Therefore Poole's definition when applied to closed default theories is equivalent to Reiter's original definition. However Example 5 below[2] shows that Poole's approach which deals only with explicitly defined theory individuals, is too weak.

Example 5. Let (D, A) be a default theory, where $D = \{\frac{:MP(x)}{P(x)}\}$, and $A = \{\exists x Q(x)\}$. Intuitively, one would expect of an implicitly defined individual satisfying Q that it satisfy P in the extension for this theory. That is, one would expect of $\exists x(P(x) \wedge Q(x))$ to belong to the extension. However, since $\bar{D} = \emptyset$, (D, A) has a unique Poole extension $E = A$.

4. Reiter's definition of open default theories

In [15] Reiter suggests an interpretation of an open default as the collection of all closed defaults of the form $\delta(\boldsymbol{t}) = \frac{\alpha(\boldsymbol{t}) : M\beta_1(\boldsymbol{t}), \ldots, M\beta_m(\boldsymbol{t})}{\gamma(\boldsymbol{t})}$, where $\boldsymbol{t} = t_1, \ldots, t_n$ is a

[2] The example is similar to that of Reiter [15, pp. 115–116], see Example 6 in the next section.

tuple of the theory individuals. However, this interpretation depends on an explicit representation of the objects under consideration, or, in other words, it depends on the underlying language. As has been pointed out in [15], when dealing with open default theories, the main problem is to specify their individuals. Reiter motivates his approach to that problem by the following examples.

Example 6 [15, pp. 115–116]. Consider an open default theory (D, A), where D consists of only one default $\frac{:MP(x)}{P(x)}$ and A contains two axioms $\exists x Q(x)$ and $\neg P(a)$. This theory contains an explicitly named individual a together with an implicitly defined individual that satisfies Q. Reiter suggests that the implicitly defined individual would satisfy P in the extension for (D, A). For this reason he introduced a new constant symbol c and replaced $\exists x Q(x)$ by $Q(c)$. This immediately yields $P(c)$ in the extension, which implies that $\exists x(Q(x) \wedge P(x))$ is also in the extension.

The next example deals with the case in which implicitly defined individuals are introduced by default.

Example 7 [15, p. 116]. Consider an open default theory (D, A), where D consists of two defaults

$$\frac{:M\exists x P(x)}{\exists x P(x)}, \qquad \frac{:MQ(x)}{Q(x)},$$

and A is empty. In this theory there is an implicitly defined individual which is introduced by the first default, and which satisfies P. One would expect, by the second default, that this individual would also satisfy Q, i.e., one would expect $\exists x(P(x) \wedge Q(x))$ to be in the extension. As in Example 6, Reiter introduces a new constant symbol c to denote this individual, and replaces the first default by $\frac{:MP(c)}{P(c)}$. This immediately yields $P(c)$ in the extension and then, by the second default, $Q(c)$ in the extension, and hence so is $\exists x(P(x) \wedge Q(x))$.

In view of these examples, when dealing with open default theories, Reiter explicitly describes the theory objects by giving names to individuals by the means of *Skolem functions*. For this purpose, he replaces the set of axioms A by its *Skolemization*, and interprets an open default by the set of closed instances of its *Skolemized form*, see the definitions below.

The *Skolemized form* of a formula ϕ is obtained as follows [16]. First put ϕ in prenex normal form ϕ'. Then replace each existentially quantified variable x of ϕ' by $f(x_1, \ldots, x_n)$, where each x_i, $i = 1, \ldots, n$, is either a free variable of ϕ', or is bound by the universal quantifier preceding $\exists x$ in the prefix of ϕ', and f is a new function symbol distinct from any in the language **L** and distinct from any other such function symbols previously introduced. Do this for all existentially quantified variables of ϕ. The result of the above transformation is a formula ϕ^{t}

without existential quantifiers and with the same free variables. Deleting all of ϕ''s quantifiers results in a quantifier-free formula ϕ^{S}, called the *Skolemization* of ϕ. The language $\mathbf{L}$ extended with the new function symbols will be denoted by $\mathbf{L}'$.

For his definition of an extension for an open default theory Reiter also needs the notion of the *Skolemized form* of a default. The Skolemized form of $\delta(x) = \frac{\alpha(x) : M\beta_1(x), \ldots, M\beta_m(x)}{\gamma(x)}$, denoted $(\delta(x))^{S}$, is a default that results from $\delta(x)$ by replacing its conclusion $\gamma(x)$ by $(\gamma(x))^{S}$. That is,

$$(\delta(x))^{S} = \frac{\alpha(x) : M\beta_1(x), \ldots, M\beta_m(x)}{(\gamma(x))^{S}} .$$

Notice that the "top part" of a default remains unchanged in converting it to its Skolemized form.

Finally, for a set of formulas A we define the *Skolemization* of A, denoted by A^{S}, by $A^{S} = \{\phi^{S} : \phi \in A\}$, for a set of defaults D we define the *Skolemization* of D, denoted by D^{S}, by $D^{S} = \{\delta^{S} : \delta \in D\}$, and the *Skolemization* of a default theory (D, A), denoted $(D, A)^{S}$, is defined by $(D, A)^{S} = (D^{S}, A^{S})$.[3]

Definition 8. Let (D, A) be an open default theory. A set of sentences E is an extension for (D, A) if and only if it is an extension for the closed default theory $(\overline{D^{S}}, A^{S})$.

Note that an extension is a set of sentences over the language $\mathbf{L}'$. Therefore, when applied to a closed default theory, Definition 4 differs from Reiter's original definition (Definition 2). In [15] Reiter suggests to overcome this problem by allowing to admit into extensions for closed default theories sentences over the language $\mathbf{L}'$.

The following examples indicate some shortcomings of Reiter's definition. In particular, when applied to close default theories, it is not equivalent to Definition 2. That is, Reiter's definition of extensions for open default theories does not pass the formal test.

Example 9. Let $A_1 = \{P(a)\}$ and $A_2 = \{P(a), \exists x P(x)\}$. Since $P(a) \vdash \exists x P(x)$, $\mathrm{Th}(A_1) = \mathrm{Th}(A_2)$. Therefore closed default theories $(\emptyset, A_1)$ and $(\emptyset, A_2)$ have the same unique extension $\mathrm{Th}(\{P(a)\})$. On the other hand, the Skolemization of $(\emptyset, A_1)$ is a closed default theory $(\emptyset, \{P(a)\})$, and the Skolemization of $(\emptyset, A_2)$ is a closed default theory $(\emptyset, \{P(a), \exists x P(x), P(c)\})$, where c is a new constant (0-place function symbol) introduced in the process of Skolemization of $\exists x P(x)$. Since without *Skolem axioms* (see Definition 12 in the next section) there is no way to deduce $P(c)$ from $P(a)$, $(\emptyset, A_1)$ and $(\emptyset, A_2)$, when considered as open default theories having different extensions.

[3] In general, the elements of A^{S} are not sentences. Therefore (D^{S}, A^{S}) does not satisfy Reiter's definition of a default theory (Definition 1). However, it is not a problem at all, because A^{S} can be equivalently replaced by $A' = \{w' : w \in A\}$

Example 10. Consider a closed default theory (D, A), where D contains two defaults

$$\frac{: MP(a)}{\forall x P(x)}, \quad \frac{: M\neg P(a)}{\neg P(a)},$$

and A consists of only one axiom $\neg P(b)$. The theory has a unique extension $\mathrm{Th}(\{\neg P(a), \neg P(b)\})$. However, its Skolemization which is an open default theory

$$(D^{S}, A^{S}) = \left(\left\{\frac{: MP(a)}{P(x)}, \frac{: M\neg P(a)}{\neg P(a)}\right\}, \{\neg P(b)\}\right)$$

has two extensions, one of which contains $\neg P(a)$ and the other contains $P(a)$.

An obvious reason for such a counterintuitive consequence of Reiter's definition is introducing the new free variable x by Skolemization of $\frac{: MP(a)}{\forall x P(x)}$. In particular, in the definition of the Skolemized form of a default, it seems more natural to replace the conclusion γ by γ^{t}.

Example 11.[4] Consider a default theory (D, A), where $D = \{\frac{: M\neg P(x)}{\neg P(x)}\}$ and $A = \emptyset$. Intuitively, the default $\frac{: M\neg P(x)}{\neg P(x)}$ expresses that $P(x)$ is assumed to be false whenever possible, and we can expect that it will allow us to prove $\forall x \neg P(x)$. But when passing to $(\overline{D^{S}}, A^{S})$, all that this default gives us is the sentences $\neg P(t)$ for the closed terms t of $\mathbf{L}'$.

In the next section we overcome the above illnesses of Reiter's definition by extending A with *Skolem axioms*, and restricting the extensions for open default theories to $\mathbf{St}_{\mathbf{L}}$, where $\mathbf{St}_{\mathbf{L}}$ denotes the set of all sentences over $\mathbf{L}$.

5. A modification to Reiter's definition of open default theories

In this section we briefly discuss a version of Reiter's approach proposed in [6]. Example 11 suggests that in order to interpret an open default as the set of its closed instances, we need the *domain closure assumption*. In [6] the domain of the theory individuals is completely described by means of Skolem functions.

Definition 12. Let f be a mapping from the set of all formulas over a language $\mathbf{L}$ of the form $\exists x \psi$ to a list of new function symbols $f_{\exists x \psi}$. We assume that f is one-to-one and if $\exists x \psi(x_1, \ldots, x_n, x)$ has exactly n free variables $x_1, \ldots, x_n$, then $f_{\exists x \psi}$ is an n-place (Skolem) function symbol. We call the language $\mathbf{L}^*$ obtained from $\mathbf{L}$ by extending the set of function symbols with $\{f_{\exists x \psi}\}$ the *Skolem expansion* of $\mathbf{L}$. The sentence

[4] This example is a simplified version of the example in [9].

$$\forall x_1 \cdots \forall x_n(\exists x\psi(x_1,\ldots,x_n,x) \supset \psi(x_1,\ldots,x_n,f_{\exists x\psi}(x_1,\ldots,x_n)))$$

is called a *Skolem axiom*. The set of all Skolem axioms is denoted by **SK**.

Finally, for a first-order theory X over $\mathbf{L}$, the *Skolem expansion* of X is a theory X^* over $\mathbf{L}^*$ defined by $X^* = X \cup \mathbf{SK}$. It is known that X^* is a *conservative expansion of* X, i.e., for an $\mathbf{L}$-sentence ϕ, $X \vdash \phi$ if and only if $X^* \vdash \phi$, see [13, Theorem 11.38(iii)(c), p. 213].

Remark 13. Skolem expansions also admit the use of the following Carnap-like rule of inference. If $X^* \vdash \phi(t)$ for each closed term of $\mathbf{L}^*$, then $X^* \vdash \forall x\phi(x)$. (Note that the Carnap rule implies the domain closure assumption.) To show that the Carnap rule is admissible in Skolem expansions, assume to the contrary that $X^* \vdash \phi(t)$ for each closed term of $\mathbf{L}^*$, but $X^* \nvdash \forall x\phi(x)$. Therefore $X^* \cup \{\exists x\neg\phi(x)\}$ has a model. Then, by [6, Lemma 1], $X^* \cup \{\neg\phi(f_{\exists x\neg\phi(x)})\}$ also has a model, in contradiction with $X^* \vdash \phi(f_{\exists x\neg\phi(x)})$.

Definition 14. A set of sentences E is a *modified Reiter extension* for (D, A) if there exists an extension E^* for $(\bar{D}, A^*)$ such that $E = E^* \cap \mathbf{St}_{\mathbf{L}}$.[5]

It was shown in [6, Theorem 3] that Definition 14, when restricted to closed default theories, is equivalent to the original definition of extension (Definition 2). That is, a set of sentences E is an extension for a closed default theory (D, A) if and only if E is a modified Reiter extension for (D, A).

Example 15. Let (D, A) be as in Example 11. That is, $D = \{\frac{:M\neg P(x)}{\neg P(x)}\}$ and $A = \emptyset$. By Lemma 3, extensions for $(\bar{D}, A^*)$ are maximal consistent sets of sentences of the form $\mathrm{Th}(\mathbf{SK} \cup A')$, where $A' \subseteq \{\neg P(t)\}_{\neg P(t)\in \mathbf{St}_{\mathbf{L}^*}}$. Since the set of sentences $\{\forall x\neg P(x)\}$ is consistent, the set of sentences $\mathbf{SK} \cup \{\forall x\neg P(x)\}$ is also consistent. Therefore $(\bar{D}, A^*)$ has a unique extension $E^* = \mathrm{Th}(\mathbf{SK} \cup \{\neg P(t)\}_{\neg P(t)\in \mathbf{St}_{\mathbf{L}^*}})$. By the Carnap rule, $\forall x\neg P(x) \in E^*$. Thus, $E^* = \mathrm{Th}(\mathbf{SK} \cup \{\forall x\neg P(x)\})$, and $E^* \cap \mathbf{St}_{\mathbf{L}} = \mathrm{Th}(\{\forall x\neg P(x)\})$ is a unique modified Reiter extension for (D, A). This extension fits the intuitive meaning of default $\frac{:M\neg P(x)}{\neg P(x)}$ saying that $P(x)$ is assumed to be false whenever possible.

The following two examples demonstrate that the above modification to Reiter's approach leads to undesirable consequences.

Example 16.[6] Let $D = \{\frac{:M\neg P(x)}{\neg P(x)}\}$ and $A = \{P(a)\}$. Similarly to Example 15, one would expect the open default theory (D, A) to have a unique modified Reiter extension $\mathrm{Th}(\{\forall x(P(x) \equiv x = a)\})$. However, as it was shown by G. Schwarz, *for every sentence ϕ such that $A \cup \{\exists x\neg P(x), \phi\}$ is consistent, (D, A) has a modified*

[5] Note that $\mathbf{SK} \vdash \phi \equiv \phi'$. Therefore there is no need to convert defaults into Skolemized form.

[6] The default theory in this example is that of the example of Lifschitz in [9].

Reiter extension that contains ϕ. Thus, in particular, (D, A) has a modified Reiter extension that contains $\neg\forall x(P(x) \equiv x = a)$. Schwarz's proof is as follows.

By Definition 14, it suffices to show that $(\bar{D}, A^*)$ has an extension containing ϕ. Let ψ denote $P(x) \wedge \neg\phi$. Then $A^* \cup \{\neg P(f_{\exists x\psi})\}$ is consistent. Indeed, let w be a model of $A \cup \{\exists x \neg P(x), \phi\}$. Then for some u in the domain of w, $w \models \neg P(u)$. Since $w \models \phi$, $w \models \neg\exists x\psi$. Therefore w can be extended to a model w^* of $A^* \cup \{\exists x \neg P(x), \phi\}$, by defining an assignment to Skolem functions in such a way that $f_{\exists x\psi}$ is assigned u, see [1, proof of Proposition 3.3.1(i), p. 164].

By Lemma 3, there is an extension E^* of $(\bar{D}, A^*)$ such that $\neg P(f_{\exists x\psi}) \in E^*$. By definition, both $P(a)$ and $\exists x(P(x) \wedge \neg\phi) \supset (P(f_{\exists x\psi}) \wedge \neg\phi)$ belong to E^*. Therefore, $E^* \vdash \neg\phi \supset P(f_{\exists x\psi})$, which together with $\neg P(f_{\in x\psi}) \in E^*$ implies $E^* \vdash \phi$. Since E^* is deductively closed, $\phi \in E^*$.

Example 17. Let (D, A) be an open default theory, where $D = \{\frac{:M\neg P(x)}{\neg P(x)}\}$, and $A = \{Q(a) \wedge P(a)\}$. This default theory (surprisingly?) has a modified Reiter extension that contains $\exists x(Q(x) \wedge \neg P(x))$. The proof is like that in Example 6. Since $Q(a) \wedge P(a) \vdash \exists x Q(x)$, by **SK**, $A^* \vdash Q(f_{\exists x Q(x)})$, and, since $A^* \nvdash P(f_{\exists x Q(x)})$, there is an extension E^* for $(\bar{D}, A^*)$ such that $\neg P(f_{\exists x Q(x)}) \in E^*$. Thus $\exists x(Q(x) \wedge \neg P(x)) \in E^* \cap \mathbf{St_L}$. But why should $Q(a) \wedge P(a)$ imply (even by default) $\exists x(Q(x) \wedge \neg P(x))$?[7]

An obvious reason for the excessive strength of the modified definition is giving explicit names for too many implicitly defined theory objects, and the *unique name assumption* for those objects. This problem can be avoided if we use Lifschitz's semantical approach that refers to the theory objects in an indirect way.

6. Semantical definition of extensions

This section contains a semantical definition of extensions for closed default theories introduced by Guerreiro and Casanova in [4], which is the starting point for Lifschitz's and our approaches in Sections 7 and 8. For what follows we need a precise definition of semantics for the first-order predicate calculus.

An **L**-*interpretation* w consists of a non-empty domain U_w, an assignment to each n-place predicate symbol P of **L** of an n-place relation P^w in U_w, and to each n-place function symbol f of **L** of an n-place function $f^w : U_w^n \to U_w$. (We treat the constants of **L** as 0-place function symbols.)

For a term $t(x_1, \ldots, x_n)$ all of whose variables are among $x_1, \ldots, x_n$, we define a function, t^w, from U_w^n to U_w by induction, as follows. Let $u_1, \ldots, u_n \in U_w$. If t is a variable x_i, then $t^w(u_1, \ldots, u_n) = u_i$; and if $t = f(t_1, \ldots, t_m)$, where f is an m-place function symbol, then

[7] A similar argument shows that Reiter's original definition suffers from the same problem.

$$t^{w}(u_1,\ldots,u_n)=f^{w}(t_1^{w}(u_1,\ldots,u_n),\ldots,t_m^{w}(u_1,\ldots,u_n))\,.$$

(Recall that we treat the constants of **L** as 0-place function symbols.) We call $t^{w}(u_1,\ldots,u_n)$ the *value of* $t(x_1,\ldots,x_n)$ *at* $u_1,\ldots,u_n$.

Let $\phi(x_1,\ldots,x_n)$ be a formula all of whose variables are among $x_1,\ldots,x_n$. We say that w *satisfies* ϕ at $u_1,\ldots,u_n$, denoted $w\models\phi(u_1,\ldots,u_n)$, if the following holds. If ϕ is an atomic formula $P(t_1,\ldots,t_n)$, then $w\models\phi$ if and only if $(t_1^{w}(u_1,\ldots,u_n),\ldots,t_n^{w}(u_1,\ldots,u_n))\in P^{w}$; $w\models\phi\supset\psi$ if and only if $w\models\phi$ implies $w\models\psi$; $w\models\neg\phi$ if and only if $w\not\models\phi$; and $w\models\forall x\phi(x)$ if and only for each $u\in U_w$, $w\models\phi(u)$.

For an interpretation w we define the **L**-*theory* of w, denoted $\mathrm{Th}_{\mathbf{L}}(w)$, as the set of all sentences of **L** satisfied by w. That is, $\mathrm{Th}_{\mathbf{L}}(w)=\{\phi\in\mathbf{St}_{\mathbf{L}}\colon w\models\phi\}$. Let X be a set of sentences over **L**. We say that w is a *model* of X, if $X\subseteq\mathbf{Th}_{\mathbf{L}}(w)$. Finally, for a class of interpretations W we define the **L**-*theory* of W, denoted $\mathrm{Th}_{\mathbf{L}}(W)$, as the set of all sentences of **L** satisfied by all the elements of W. That is, $\mathrm{Th}_{\mathbf{L}}(W)=\bigcap_{w\in W}\mathrm{Th}_{\mathbf{L}}(w)$.

We say that interpretations w_1 and w_2 are **L**-*equivalent*, if $\mathrm{Th}_{\mathbf{L}}(w_1)=\mathrm{Th}_{\mathbf{L}}(w_2)$.

Note that the assignment to the equality relation in interpretations is a binary relation that does not have to be identity in the domain of the interpretation, but satisfies the equality first-order axioms. (Interpretations where the assignment to the equality relation is identity in the domain of the interpretation are called *normal*, see [12, p. 78] for details.) That is, equality relation is treated as an ordinary dyadic predicate which satisfies the equality axioms.

Extensions for a closed default theory can be defined semantically as follows.

Definition 18 [4]. Let (D, A) be a closed default theory. For any class of interpretations W let $\Sigma_{(D,A)}(W)$ be the largest class of models of A that satisfies the following condition.

If $\frac{\alpha\,:\,M\beta_1,\ldots,M\beta_m}{\gamma}\in D$, $\alpha\in\mathrm{Th}_{\mathbf{L}}(\Sigma_{(D,A)}(W))$, and $\neg\beta_1,\ldots,\neg\beta_m\not\in\mathrm{Th}_{\mathbf{L}}(W)$, then $\gamma\in\mathrm{Th}_{\mathbf{L}}(\Sigma_{(D,A)}(W))$.[8]

It is known from [4] that the definition of extensions as the theories of the fixed points of Σ is equivalent to Reiter's original definition (Definition 2). That is, a set of sentences E is an extension for a closed default theory (D, A) if and only if $E=\mathrm{Th}_{\mathbf{L}}(W)$ for some fixed point W of $\Sigma_{(D,A)}$.

7. Lifschitz's approach to open default theories.

In this section we examine Lifschitz's definition of extensions for open default theories, called a *default logic with a fixed universe* [9, Section 3]. This definition involves a class of interpretations, called U-worlds, which are defined below.

[8] This largest class $\Sigma^{U}_{(D,A)}(W)$ always exists, see [9, Proposition 1].

Definition 19. Let U be a non-empty set. A U-world is a normal **L**-interpretation w such that $U_w = U$. That is, the equality relation of **L** is interpreted in w by identity: $w \models u_1 =^w u_2$ if and only if u_1 and u_2 are the same element of U.

Lifschitz's definition of extensions for open default theories is a relativization of Definition 18 to U-worlds.

Definition 20. Let U be a non-empty set and let (D, A) be a default theory. For any set of U-worlds W let $\Delta^U_{(D,A)}(W)$ be the largest set, V, of U-worlds which are models of A and for which the following condition is satisfied.

For any $\frac{\alpha(x) : M\beta_1(x), \ldots, M\beta_m(x)}{\gamma(x)} \in D$, and any tuple $\boldsymbol{u}$ of elements of U, if $\alpha(\boldsymbol{u}) \in \mathrm{Th}_{\mathbf{L}_U}(V)$, and $\neg\beta_1(\boldsymbol{u}), \ldots, \neg\beta_m(\boldsymbol{u}) \not\in \mathrm{Th}_{\mathbf{L}_U}(W)$, then $\gamma(\boldsymbol{u}) \in \mathrm{Th}_{\mathbf{L}_U}(V)$, where $\mathbf{L}_U$ denotes the language obtained from **L** by extending its set of constants with all elements of U.[9]

We refer to the set W as the set of *possible* worlds, and we refer to the set V as the set of *belief* worlds. Note that the sentences of $\mathbf{L}_U$ are of the form $\phi(u_1, \ldots, u_n)$, where $u_1, \ldots, u_n \in U$, and $\phi(x_1, \ldots, x_n)$ is a formula of **L** all of whose free variables are among $x_1, \ldots, x_n$.

A set of sentences E is called a *U-extension* for (D, A) if $E = \mathrm{Th}_{\mathbf{L}}(W)$ for some fixed point W of $\Delta^U_{(D,A)}$. Below U-extensions are referred to as *Lifschitz extensions*.

As it was pointed out in [9], Definition 20 when applied to closed default theories is not equivalent to Definition 2, because the cardinality of U can be extracted from a U-extension in the following manner. For a positive integer n consider the sentence

$$\exists x_1 \cdots \exists x_n \Big(\bigwedge_{1 \leq i < j \leq n} x_i \neq x_j \wedge \forall x \Big(\bigvee_{1 \leq i \leq n} x = x_i \Big) \Big),$$

denoted e_n, which states that there are exactly n distinct theory objects. For a domain U we defined $card_U = \{e_n\}$, if U is of a finite cardinality n, and $card_U = \{\neg e_n\}_{n=1,2,\ldots}$, if U is infinite. It immediately follows from Definitions 19 and 20 that any U-extension must contain $card_U$ as a subset, *even for closed default theories* (whose extensions do not depend on U). Thus Lifschitz's approach does not pass the formal test.[10]

As we shall see in the next section, the above deficiency of Lifschitz's approach can be eliminated, if we replace U-worlds by infinite non-normal interpretations with the same domain. That is, the equality symbol does not have to be interpreted by identity, but by an equivalence relation that satisfies the equality axioms, exactly as in the classical model theory. (For example e_n has a model w with an infinite domain U_w. In this model the number of the equivalence classes of U_w modulo $=^w$ is n.) However, Lifschitz's definition has a deeper intrinsic problem, illustrated by the following example.

[9] This largest set $\Delta^U_{(D,A)}(W)$ always exists, see [9, Proposition 3].

[10] Example 21 below shows that it is also counterintuitive.

Example 21. Consider an open default theory (D, A), where $D = \{\frac{:M\neg P(x)}{\neg P(x)}\}$ and $A = \{a \neq b, \forall x(x = a \vee x = b), \exists x P(x), \exists x \neg P(x)\}$. The axioms $\forall x(x = a \vee x = b)$ and $a \neq b$ are an instance of the *domain closure assumption* stating that each theory object is represented either by a or by b. Thus, (D, A) has a U-extension only when the cardinality of U is 2, which avoids the above deficiency of Lifschitz's definition in the case of this example, but raises a new problem.

Given $\forall x(x = a \vee x = b)$, it seems very natural to replace D by the set of its all closed instances

$$D^{c} = \left\{ \frac{:M\neg P(a)}{\neg P(a)}, \frac{:M\neg P(b)}{\neg P(b)} \right\},$$ [11]

and we would expect the set of extensions for an open default theory (D, A) to coincide with the set of extensions for a closed default theory (D^{c}, A). It immediately follows from Lemma 3 that (D^{c}, A) has two extensions $\mathrm{Th}(A \cup \{\neg P(a)\})$ and $\mathrm{Th}(A \cup \{\neg P(b)\})$.

However (D, A) has only one Lifschitz extension $\mathrm{Th}(A)$. Indeed, let $U = \{u_1, u_2\}$. It can be easily verified that $\Delta^{U}_{(D,A)}$ has two fixed points $V = \{v_1, v_2\}$ and $W = \{w_1, w_2\}$ which are defined by the following table.

	v_1	v_2	w_1	w_2
a	u_1	u_2	u_1	u_2
b	u_2	u_1	u_2	u_1
P	$\{u_1\}$	$\{u_1\}$	$\{u_2\}$	$\{u_2\}$

In this table, the meaning of the column marked v_1 is that the assignment to constants in v_1 is defined by $a^{v_1} = u_1$ and $b^{v_1} = u_2$, and the assignment to P in v_1 is defined by $P^{v_1} = \{u_1\}$. The meaning of the other columns is similar.

Even though v_1 and v_2 (w_1 and w_2) belong to the same fixed point, their assignments to a and b are different. Therefore neither $P(a)$, nor $P(b)$ belongs to $\mathrm{Th}_{\mathrm{L}}(V)$ ($\mathrm{Th}_{\mathrm{L}}(W)$). Thus $\mathrm{Th}_{\mathrm{L}}(V) = \mathrm{Th}_{\mathrm{L}}(W) = \mathrm{Th}(A)$.

Remark 22. In [9] Lifschitz claims that his formalization of default reasoning does not assume the domain closure assumption. However, for each fixed point W of Δ^{U} the following holds. If $\phi(u) \in \mathrm{Th}_{\mathrm{L}_U}(W)$ for every $u \in U$, then $\forall x \phi(x) \in \mathrm{Th}_{\mathrm{L}_U}(W)$ as well. In other words, the Carnap rule (which is equivalent to the domain closure assumption) is admissible for $\mathrm{Th}_{\mathrm{L}_U}(W)$. In general, it is unclear how to deal with open defaults without the closure domain assumption, because we need a way to specify all objects of the theory, see [7] for a general discussion. Moreover, the domain closure assumption together with the fact that in a U-world

[11] To some extent, such a replacement is analogous to the completion of a program in logic programming.

the assignment to the equality of **L** is identity (which corresponds to the unique name assumption, see Proposition 43 in the next section) is the reason for which we can extract the cardinality of U from a U-extension.

One of the reasons for counterintuitive consequences of Lifschitz's approach is that it does not distinguish between the assignments to explicitly and implicitly defined function symbols. In particular, in Example 21, the 0-place function symbols (constants) a and b are treated as objects implicitly defined by their properties. The precise definition is as follows.

Definition 23. Let P be a mapping from the set of all function symbols f of a language **L** to a list of new predicate symbols P_f. We assume that P is one-to-one and, if f is an n-place function symbol, then P_f is an $(n+1)$-place (defining) predicate symbol. We call the language $\mathbf{L}_\mathrm{D}$ obtained from **L** by deleting its function symbols and extending its predicate symbols with P_f's the *defining expansion* of **L**. The sentence $\forall x_1 \cdots \forall x_n \exists! x P_f(x_1, \ldots, x_n, x)$ is called a *defining axiom*. The set of all defining axioms is denoted by **D**.

For a formula ϕ of **L** we define its translation, denoted ϕ^P, into $\mathbf{L}_\mathrm{D}$, by induction, as follows. Consider a sequence $\phi_0, \phi_1, \ldots$ of formulas over the language obtained by extending **L** with the new predicate symbols, such that ϕ_0 is ϕ, and ϕ_{i+1} results from ϕ_i in the following manner. If ϕ_i does not contain function symbols, then the sequence terminates at ϕ_i. Otherwise, let $f(t_1, \ldots, t_n)$ be the leftmost term in ϕ_i such that all t_i are variables, and let $P(\ldots, f(t_1, \ldots, t_n), \ldots)$ be the atomic subformula of ϕ_i that contains that term. Then ϕ_{i+1} is obtained from ϕ_i by replacing $P(\ldots, f(t_1, \ldots, t_n), \ldots)$ with $\exists x(P_f(t_1, \ldots, t_n, x) \wedge P(\ldots, x, \ldots))$. Since ϕ_{i+1} contains one function symbol less than ϕ_i, the above sequence must terminate. The last formula in the sequence is ϕ^P. For a set of **L**-formulas X we define a set of $\mathbf{L}_\mathrm{D}$-formulas X^P by $X^\mathrm{P} = \{\phi^\mathrm{P} : \phi \in X\}$, and for a set D of defaults over **L** we define a set D^P of defaults over $\mathbf{L}_\mathrm{D}$ by

$$D^\mathrm{P} = \left\{ \frac{\alpha^\mathrm{P}(\boldsymbol{x}) : M\beta_1^\mathrm{P}(\boldsymbol{x}), \ldots, M\beta_m^\mathrm{P}(\boldsymbol{x})}{\gamma^\mathrm{P}(\boldsymbol{x})} : \frac{\alpha(\boldsymbol{x}) : M\beta_1(\boldsymbol{x}), \ldots, M\beta_m(\boldsymbol{x})}{\gamma(\boldsymbol{x})} \in D \right\}.$$

Finally, for a first-order theory X over **L**, the *defining expansion* of X is a theory X^D over $\mathbf{L}_\mathrm{D}$ defined by $X^\mathrm{D} = X^\mathrm{P} \cup \mathbf{D}$.

Proposition 24 below establishes a relationship between Lifschitz extensions for a default theory and its defining expansion, and sheds more light on the reason for the counterintuitive behavior of Lifschitz extensions demonstrated by Example 10.

Proposition 24. *Let (D, A) be a default theory. Then E is a Lifschitz extension for (D, A) if and only if E^P is a Lifschitz extension for $(D^\mathrm{P}, A^\mathrm{D})$.*

Proof. Let w be an **L**-interpretation. Consider an $\mathbf{L}_\mathrm{D}$-interpretation w^P such that

$U_{w^{\mathrm{P}}} = U_w$, for any predicate symbol P of $\mathbf{L}$, $P^{w^{\mathrm{P}}} = P^w$, and for any n-place function symbol f of $\mathbf{L}$, the assignment to P_f is defined by

$$P_f^{w^{\mathrm{P}}} = \{(u_1, \ldots, u_n, f^w(u_1, \ldots, u_n)) \colon u_1, \ldots, u_n \in U_w\}.$$

It can be readily verified that $w^{\mathrm{P}} \models \mathbf{D}$, and that for any $\mathbf{L}_{U_w}^{\mathrm{D}}$ sentence $\phi(u_1, \ldots, u_n)$, $w \models \phi(u_1, \ldots, u_n)$ if and only if $w^{\mathrm{P}} \models \phi^{\mathrm{P}}(u_1, \ldots, u_n)$.

Conversely, let w be an $\mathbf{L}_{\mathrm{D}}$-interpretation such that $w \models \mathbf{D}$. Consider an $\mathbf{L}$-interpretation w^{F} such that $U_{w^{\mathrm{F}}} = U_w$, for any predicate symbol P of $\mathbf{L}$, $P^{w^{\mathrm{F}}} = P^w$, and for any n-place function symbol f of $\mathbf{L}$ and $u_1, \ldots, u_n \in U_w$, $f^{w^{\mathrm{F}}}(u_1, \ldots, u_n)$ is defined as follows. By $\mathbf{D}$, there is a unique (modulo $=^w$) $u \in U$ such that $w \models P_f^w(u_1, \ldots, u_n, u)$, and we put $f^{w^{\mathrm{F}}}(u_1, \ldots, u_n) = u$. Again, for any $\mathbf{L}_{U_w}$ sentence $\phi(u_1, \ldots, u_n)$, $w^{\mathrm{F}} \models \phi(u_1, \ldots, u_n)$ if any only if $w \models \phi^{\mathrm{P}}(u_1, \ldots, u_n)$.

Now it follows that W is a fixed point of $\Delta^U_{(D,A)}$ if and only if $W^{\mathrm{P}} = \{w^{\mathrm{P}} \colon w \in W\}$ is a fixed point of $\Delta^U_{(D^{\mathrm{P}}, A^{\mathrm{D}})}$, which completes the proof. □

8. A modification of Lifschitz's approach

In this section we modify Lifschitz's approach to make it free from the deficiencies discussed in the previous section. An obvious way to avoid the reference to the cardinality of the interpretation is to replace U-worlds by ordinary interpretations, and avoiding the function (constant) interpretation problem in Example 21, can be achieved by "separating" the assignments to functions and predicate symbols in an interpretation. In particular, we should require the interpretations under consideration to have the same assignments to function symbols. Example 29 below formalizes these stronger requirements, but shows that they are still insufficient. First we need the following definition.

Definition 25. An $\mathbf{L}$-*pre-interpretation* F consists of a non-empty domain U_F and an assignment to each n-place function symbol f of $\mathbf{L}$, of an n-place function $f^F : U_F^n \to U_F$. (Recall that we treat the constants of $\mathbf{L}$ as 0-place function symbols.)

In view of the above definition, an interpretation can be thought of as a pre-interpretation together with an assignment to the predicate symbols, or, alternatively, a pre-interpretation can be thought as a "function structure" of corresponding interpretations.[12] The pre-interpretation of an interpretation w is denoted by F_w, and we say that w is *based* on F_w.

[12] Pre-interpretations are widely used in logic programming to separate between the assignments to function and predicate symbols, see [10]. Usually only one pre-interpretation is considered, the free, Herbrand one. Our modification of Lifschitz's approach is also based on free pre-interpretations, see Definition 34 below.

Definition 26. An L-interpretation based on a pre-interpretation F is called an *F-world*.

The opening paragraph of this section suggests to modify Definition 20 by adding the following *compatibility* requirement.

Definition 27. Let F be a pre-interpretation and let (D, A) be a default theory. For any set W of F-worlds let $\Delta^F_{(D,A)}(W)$ be the largest set, V, of F-worlds which are models of A that satisfies the following condition.

For any $\frac{\alpha(x) : M\beta_1(x), \ldots, M\beta_m(x)}{\gamma(x)} \in D$, and any tuple $\boldsymbol{u}$ of elements of U_F, if $\alpha(\boldsymbol{u}) \in \mathrm{Th}_{\mathbf{L}_{U_F}}(V)$, and $\neg\beta_1(\boldsymbol{u}), \ldots, \neg\beta_m(\boldsymbol{u}) \not\in \mathrm{Th}_{\mathbf{L}_{U_F}}(W)$, then $\gamma(\boldsymbol{u}) \in \mathrm{Th}_{\mathbf{L}_{U_F}}(V)$.[13]

A set of sentences E is called an F-extension for (D, A) if there is a fixed point W of $\Delta^F_{(D,A)}$ such that $E = \mathrm{Th}_{\mathbf{L}}(W)$. (Note that W must be a set of F-worlds.)

Remark 28. The notion of F-extensions is similar to that of *default logic with fixed constants* [9, Section 6], also see Definition 49 in the next section. In Lifschitz's terminology, F-extensions are extensions all of whose functions symbols are fixed.

Example 29. Consider a *closed* default theory $(\{\frac{: Ma \neq b}{a \neq b}\}, \emptyset)$ which has a unique extension $\mathrm{Th}(\{\alpha \neq b\})$. However it has two F-extensions: $\mathrm{Th}(\{a = b\})$ and $\mathrm{Th}(\{a \neq b\})$. The former corresponds to the pre-interpretations with the same assignment to a and b, and the letter corresponds to the pre-interpretations, where the assignments to a and b are different.

Example 29 shows that the notion of compatibility introduced by Definitions 25–27 is not sufficient for the following reason. Fixing the assignments to function symbols might impose some undesirable constraints on the assignment to equality in the interpretation. (In the example, $a^w =^w b^w$ is such a constraint on the assignment to equality in the interpretation w.) Therefore, if we want to proceed in that way, we must restrict ourselves to *independent* assignments to function symbols, i.e., to pre-interpretations which, roughly speaking, assign distinct elements of the domain to different terms.

Definition 30. Let b be a set that contains no symbols of $\mathbf{L}$. The *Herbrand pre-interpretation* for $\mathbf{L}_b$,[14] denoted H_b, is the pre-interpretation whose domain U_{H_b} consist of all closed terms of $\mathbf{L}_b$, and for any n-place function symbol f and closed terms $t_1, \ldots, t_n$ of $\mathbf{L}_b$, $f^{H_b}(t_1, \ldots, t_n) = f(t_1, \ldots, t_n)$. The set b is called the *base* of H_b.[15]

[13] Similarly to [9, Proposition 3] it can be shown that this largest class $\Delta^F_{(D,A)}(W)$ always exists. In fact, a routine inspection of the proof shows that it holds in the case of pre-interpretations as well.

[14] Recall that $\mathbf{L}_b$ is obtained from $\mathbf{L}$ by extending the set of constants of $\mathbf{L}$ with the elements of b.

[15] It can be shown that any pre-interpretation can be obtained (up to an isomorphism) by factorizing a Herbrand pre-interpretation modulo some congruence relation, see the proof of Proposition 33.

Definition 31. A pre-interpretation is called *free* if it is isomorphic to a Herbrand pre-interpretation. The base b_F of a free pre-interpretation F is the subset of its domain U_F which consists of those elements u such that for no (n-place) function symbol f of **L** and for no $u_1, \ldots, u_n \in U_F$, $u = f^F(u_1, \ldots, u_n)$.

Free pre-interpretations play a major role in our modification to Lifschitz's approach; for a better intuition, some of their basic properties are stated below.

For any free pre-interpretation F the following holds.

- For all pairs of m- and n-place distinct function symbols g and f of **L** and for all $u_1, \ldots, u_m, v_1, \ldots, v_n \in U_F$, $g^F(u_1, \ldots, u_m) \neq f^F(v_1, \ldots, v_n)$. (Recall that we treat the constants of **L** as 0-place function symbols.)
- For each n-place function symbol f of **L** and for all $u_1, \ldots, u_n, v_1, \ldots, v_n \in U_F$ such that for some $i = 1, \ldots, n$, $u_i \neq v_i$, $f^F(u_1, \ldots, u_n) \neq f^F(v_1, \ldots, v_n)$.
- For each term $t(x_1, \ldots, x_n)$ of **L**, whose set of free variables is $\{x_1, \ldots, x_n\}$, and for all $u_1, \ldots, u_n \in U_F$, $t^F(u_1, \ldots, u_n) \neq u_i$, $i = 1, \ldots, n$. Moreover, the following binary relation $<_F$ on U_F is well-founded. For $u,v \in U_F$, $u <_F v$ if and only if there exist an n-place function symbol f of **L**, $n > 0$, and $u_1, \ldots, u_n \in U_F$ such that for some $i = 1, \ldots, n$, $u_i = u$ and $f^F(u_1, \ldots, u_n) = v$.

The above properties of free pre-interpretations show that normal interpretations based on free pre-interpretations satisfy the *equality theory*, see [10, p. 79] for the axioms of the equality theory.[16] Conversely, it can be shown that if a pre-interpretation of a model of the equality theory is well-founded, then it is free.

Definition 32. A free pre-interpretation with an infinite base is called *universal*.[17]

The "universal" property of universal pre-interpretations is given by Proposition 33 below.

Proposition 33. *Let A be a consistent* **L**-*theory, and let F be a universal pre-interpretation. Then A has a model whose pre-interpretation is F.*

Proof. Let b be the base of F. Renaming the elements of U_F, if necessary, we may assume that $F = H_b$. Since b is infinite, there is a model w of A such that $U_w = b$, see [1, Corollary 2.1.6, p. 67]. For each $u \in U_{H_b}$, we define an element $\varepsilon(u)$ of $b(=U_w)$, by $\varepsilon(u) = u$ for $u \in b$, and for a closed term $f(t_1, \ldots, t_n)$ of $\mathbf{L}_b$ we define $\varepsilon(f(t_1, \ldots, t_n))$ by $\varepsilon(f(t_1, \ldots, t_n)) = f^w(\varepsilon(t_1), \ldots, \varepsilon(t_n))$.

---

[16] In logic programming the equality theory is used in the context of defining negation by Clark's *completion*. In particular, it allows a consistency between equality in a model and the result of unification between first-order terms, see [10] for a detailed exposition of the subject.

[17] Universal pre-interpretations naturally arise in the Henkin proof of the completeness theorem, see [12, Lemma 2.16].

Consider an interpretation w_b such that $F_{w_b} = H_b$ and for an n-place predicate symbol P, $P^{w_b} = \{(u_1, \ldots, u_n): (\varepsilon(u_1), \ldots, \varepsilon(u_n)) \in P^w\}$. We prove by induction on the length of an **L**-formula $\phi(x_1, \ldots, x_n)$ that $w_b \models \phi(u_1, \ldots, u_n)$ if and only if $w \models \phi(\varepsilon(u_1), \ldots, \varepsilon(u_n))$. By the definition of w_b this is true for atomic formulas, and the case of propositional connectives $\supset$ and $\neg$ is immediate. For the case of the existential quantifier, if $w_b \models \exists x\phi(x, u_1, \ldots, u_n)$, then for some $u \in U_{w_b}$, $w_b \models \phi(u, u_1, \ldots, u_n)$. By the induction hypothesis, $w \models \phi(\varepsilon(u), \varepsilon(u_1), \ldots, \varepsilon(u_n))$, which implies $w \models \exists x\phi(x, \varepsilon(u_1), \ldots, \varepsilon(u_n))$. Conversely, if $w \models \exists x\phi(x, \varepsilon(u_1), \ldots, \varepsilon(u_n))$, then for some $u \in U_w$, $w \models \phi(u, \varepsilon(u_1), \ldots, \varepsilon(u_n))$. By definition, $u = \varepsilon(u)$. Thus, by the induction hypothesis, $w_b \models \phi(u, u_1, \ldots, u_n)$, which implies $w_b \models \exists x\phi(x, u_1, \ldots, u_n)$. □

Our modification to Lifschitz's approach (Definitions 19 and 20) is replacing U-worlds by F-worlds, for a universal pre-interpretation F.

Definition 34. Let F be a universal pre-interpretation and let (D, A) be a default theory. We call F-extensions for (D, A) *modified Lifschitz extensions* or *universal extensions*.

The difference between Definition 34 and Lifschitz's original definition (Definition 20) is the replacement of sets of normal interpretations over the same possibly finite domain by sets of interpretations based on the same universal pre-interpretation. (In other words we allow the assignments to equality to *vary* over the pre-interpretation domain.) In particular, whereas Lifschitz's original definition, requiring the normal interpretation of the equality relation, emphasizes the *predicate structure* of the interpretation, our approach is based on the compatibility of the function assignments.

Examples 38–41 below (which are continuations of Examples 5, 16, 17, and 21 respectively) show the behavior of the modified definition in some cases where both Reiter's and Lifschitz's approaches are counterintuitive. The proofs in this section are based on the following definition and lemmas. In particular, Lemma 37 is similar to Lemma 3 and is a relativization of [9, Proposition 6] to the F-worlds.

Definition 35. Let $\beta(\boldsymbol{x})$ be an **L**-formula. For an interpretation w, β^w denotes the set of tuples $\boldsymbol{u}$ such that $w \models \beta(\boldsymbol{u})$. Let D be a set of normal defaults without prerequisites and let F be a pre-interpretation. We say that an F-world w is *D-maximal* if there is no F-world w' such that for all $\frac{:M\beta(\boldsymbol{x})}{\beta(\boldsymbol{x})} \in D$, $\beta^w \subseteq \beta^{w'}$; and for some $\frac{:M\beta(\boldsymbol{x})}{\beta(\boldsymbol{x})} \in D$, β^w is a proper subset of $\beta^{w'}$. Finally, we say that F-worlds w_1 and w_2 are *D-equivalent*, denoted $w_1 \sim_D w_2$, if $\beta^{w_1} = \beta^{w_2}$, for all $\frac{:M\beta(\boldsymbol{x})}{\beta(\boldsymbol{x})} \in D$. (It immediately follows from the definition that $\sim_D$ is an equivalence relation.)

Lemma 36. *Let (D, A) be a default theory and let D contain a normal default without prerequisites $\frac{:M\beta(\boldsymbol{x})}{\beta(\boldsymbol{x})}$. Let F be an* **L**-*pre-interpretation and let W be a set of*

F-worlds. If for some $w \in W$ and some tuple $\boldsymbol{u}$ of elements of U_F, $w \models \beta(\boldsymbol{u})$, then for any $v \in \Delta^F_{(D,A)}(W)$, $v \models \beta(\boldsymbol{u})$.

Proof. The proof follows immediately from the definition of $\Delta^F_{(D,A)}$ (Definition 27). $\square$

The following lemma is similar to [9, Proposition 6] which shows that the fixed points of $\Delta^U_{(\{\frac{:M\beta(x)}{\beta(x)}\},A)}$ correspond to the extends of β in $\{\beta\}$-maximal U-worlds.

Lemma 37. *Let F be a pre-interpretation and let D be a set of normal defaults without prerequisites. A set of F-worlds is a fixed point of $\Delta^F_{(D,A)}$ if and only if it is an equivalence class of the restriction of $\sim_D$ to the D-maximal F-worlds which are models of A.*

Proof. Let W be an equivalence class of $\sim_D$ restricted to the D-maximal F-worlds which are models of A. Then the set of belief F-worlds W satisfies the condition of Definition 27 for the set of possible worlds W. In order to show that W is the maximal set of F-worlds satisfying that condition assume to the contrary, that it is a proper subset of $\Delta^F_{(D,A)}(W)$. By definition, $\Delta^F_{(D,A)}(W) \models A$, and, by Lemma 36, each element of $\Delta^F_{(D,A)}(W)$ is D-equivalent to the elements of W, and, therefore, is D-maximal. Since W is an equivalence class of $\sim_D$ restricted to the D-maximal F-worlds which are models of A, $\Delta^F_{(D,A)}(W) \subseteq W$, in contradiction with the assumption that W is a proper subset of $\Delta^F_{(D,A)}(W)$. This proves the "if" direction of the lemma.

Conversely, let W be a fixed point of $\Delta^F_{(D,A)}$. By Lemma 36, for any two worlds $w_1, w_2 \in W$ and for all $\frac{:M\beta(x)}{\beta(x)} \in D$, $\beta^{w_1} = \beta^{w_2}$. Therefore, for all $\frac{:M\beta(x)}{\beta(x)} \in D$ we may write β^W for β^w, where $w \in W$. In order to prove that W is an equivalence class of the restriction of $\sim_D$ to the D-maximal F-worlds which are models of A, assume to the contrary that there exists an F-world $w' \not\in W$ such that $w' \models A$ and for all $\frac{:M\beta(x)}{\beta(x)} \in D$, $\beta^W \subseteq \beta^{w'}$. Then the set of belief worlds $V = W \cup \{w'\}$ satisfies the condition of Definition 27 for the default theory (D, A) and the set of possible worlds W, in contradiction with the maximality of W. This proves the "only if" direction of the lemma. $\square$

Example 38. Let (D, A) be as in Example 5. That is, $D = \{\frac{:MP(x)}{P(x)}\}$, and $A = \{\exists x Q(x)\}$. Obviously, if an F-world w satisfying $\exists x Q(x)$ is P-maximal, then $w \models \forall x P(x)$. Therefore (D, A) has a unique modified Lifschitz extension $\{\forall x P(x), \exists x Q(x)\}$. In particular, this extension contains $\exists x(P(x) \wedge Q(x))$, which is what we expect in our case of an implicitly defined individuals, see the discussion in Example 5.

Example 39. Let (D, A) be as in Example 16. That is, $D = \{\frac{:M\neg P(x)}{\neg P(x)}\}$, and $A = \{P(a)\}$. Let w be a $\neg P$-maximal F-world. Then $(\neg P)^w = U_w - \{a^w\}$. Thus, by Lemma 37 (D, A) has a unique modified Lifschitz extension $\mathrm{Th}(\{\forall x(P(x) \equiv x =$

$a)\})$. We see that, like in Example 15, the meaning of default $\frac{:M\neg P(x)}{\neg P(x)}$ is that $P(x)$ is assumed to be false whenever possible.

Example 40. Let (D, A) be as in Example 17. That is, $D = \{\frac{:M\neg P(x)}{\neg P(x)}\}$, and $A = \{Q(a) \wedge P(a)\}$. Let w be a $\neg P$-maximal F-world. Then $(\neg P)^w = U_w - \{a^w\}$. Thus, by Lemma 37, (D, A) has a unique modified Lifschitz extension $\mathrm{Th}(\{\forall x((Q(x) \wedge P(x)) \equiv x = a)\})$. That is, as above, $P(x)$ is assumed to be false whenever possible.

Example 41. Let (D, A) be as in Example 21. That is, $D = \{\frac{:M\neg P(x)}{\neg P(x)}\}$ and $A = \{a \neq b, \forall x(x = a \vee x = b), \exists x P(x), \exists x \neg P(x)\}$. By Lemma 37, (D, A) has two modified Lifschitz extensions $\mathrm{Th}(A \cup \{\neg P(a)\})$ and $\mathrm{Th}(A \cup \{\neg P(b)\})$, which are exactly what we expect in our case of an explicitly defined finite domain, see the discussion in Example 21.

The next theorem shows that modified Lifschitz extensions pass the formal test. That is, for closed default theories they coincide with the ordinary extensions.

Theorem 42. *Let (D, A) be a closed default theory. Then a set of sentences is an extension for (D, A) if and only if it is a modified Lifschitz extension for (D, A).*

Proof. Let E be an extension for (D, A) and let F be a universal $\mathbf{L}^*$-pre-interpretation. By Proposition 33, for each model w of E there is an L-equivalent interpretation w_F based on F.

Let $W = \{w_F : w \models E\}$. By the definition of w_F, W is a set of F-worlds such that $E = \mathrm{Th}_{\mathbf{L}}(W)$. Thus in order to prove that E is an F-extension for (D, A), it suffices to show that W is the largest set of belief F-worlds that satisfies the condition of Definition 27 for the default theory (D, A) and the set of possible worlds W.

Let $\frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma} \in D$, such that $\alpha \in \mathrm{Th}_{\mathbf{L}_{U_F}}(W)$, and $\neg\beta_1, \ldots, \neg\beta_m \not\in \mathrm{Th}_{\mathbf{L}_{U_F}}(W)$. (Recall that (D, A) is a closed default theory.) Since, $\mathrm{Th}_{\mathbf{L}}(W) = E$ and E is an extension for (D, A), $\gamma \in E$ $(=\mathrm{Th}_{\mathbf{L}}(W))$, i.e., the condition of Definition 27 is satisfied. To prove that W is the largest set satisfying that condition, assume to the contrary, that $\Delta^F_{(D,A)}(W)$ contains W as a proper subset. Let $w_F \in \Delta^F_{(D,A)}(W) - W$. Then, by the definition of W, $w_F \not\models E$, which implies that $B = \mathrm{Th}(\Delta^F_{(D,A)}(W))$ is a proper subset of E. Obviously, B satisfies properties (D1)–(D3) of Definition 2 for (D, A) and $S = E$, in contradiction with the minimality of E. This completes the proof of the "only if" part of the theorem.

Conversely, let E be an F-extension for (D, A). That is $E = \mathrm{Th}_{\mathbf{L}}(\Delta^F_{(D,A)}(W))$, for some fixed point W of $\Delta^F_{(D,A)}$. Obviously, $B = E$ satisfies properties (D1)–(D3) of Definition 2 for (D, A) and $S = E$. It remains to prove that B is the minimal set satisfying those properties. Assume to the contrary, that $\Gamma_{(D,A)}(E)$ is a proper subset of $B(=E)$. Then, by Proposition 33, there is a model w of $\Gamma_{(D,A)}(E)$ such that $F_w = F$, and $w \not\models E$. Thus $w \not\in W$. It follows that the set of belief worlds

$W \cup \{w\}$ satisfies the condition of Definition 27 for the default theory (D, A) and the set of possible worlds W, in contradiction with the maximality of W, which completes the proof of the theorem. □

Propositions 43 and 44 below give a relationship between the original and modified Lifschitz's approaches. They state the Lifschitz extensions for a rather large class of default theories can be defined in terms of modified Lifschitz extensions. Proposition 43 shows that the cardinality of domain U can be expressed by the axiom $card_U$, and the reference to normal interpretations can be expressed by the default $\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}$ which corresponds to the unique name assumption, see [7, Section 3]. In Proposition 44 we simulate the unique name assumption by two defaults:

$$\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}, \quad \frac{: Mx_1 = x_2}{x_1 = x_2},$$ [18]

which allows to describe Lifschitz extension in terms of modified Lifschitz extensions without explicit reference to the cardinality of the domain.

Proposition 43. *Assume that* L *contains no function symbols. Then a set of sentences is a Lifschitz extension for* (D, A) *if and only if for some* U, *it is a modified Lifschitz extension for*

$$\left(D \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\}, A \cup card_U\right).$$

Proposition 44. *Assume that* L *contains no function symbols. Then a set of sentences is a Lifschitz extension for* (D, A) *if and only if it is a modified Lifschitz extension for*

$$\left(D \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}, \frac{: Mx_1 = x_2}{x_1 = x_2}\right\}, A\right).$$

Propositions 43 and 44 are immediate corollaries, respectively, of Theorems 53 and 54 which we prove in the next section. The theorems generalize the propositions to default logics with fixed constants.

The following corollaries to Propositions 43 and 44 show how Lifschitz extensions can be translated into modified ones.

Corollary 45. *Let* (D, A) *be a default theory. Then a set of sentences is a Lifschitz extension for* (D, A) *if and only if for some* U, E^{P} *is a modified Lifschitz extension for*

[18] These defaults correspond to the modal *strict equality* axioms $M(x_1 \neq x_2) \supset (x_1 \neq x_2)$ and $M(x_1 = x_2) \supset (x_1 = x_2)$, see [5, p. 190].

$$\left(D^{\mathrm{P}} \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\}, A^{\mathrm{D}} \cup card_U\right).$$

Proof. The proof follows immediately from Propositions 24 and 43. □

Corollary 46. *Let (D, A) be a default theory. Then a set of sentences is a Lifschitz extension for (D, A) if and only if E^{P} is a modified Lifschitz extension for*

$$\left(D^{\mathrm{P}} \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}, \frac{: Mx_1 = x_2}{x_1 = x_2}\right\}, A^{\mathrm{D}}\right).$$

Proof. The proof follows immediately from Propositions 24 and 44. □

9. Default logic with fixed constants

In this section we express extensions for *default logics with fixed constants* introduced in [9, Section 6] in terms of modified Lifschitz extensions, see Theorems 53 and 54 below. These theorems show that a predicate P can be fixed by two "opposite" defaults $\frac{: M\neg P(x)}{\neg P(x)}$ and $\frac{: MP(x)}{P(x)}$. That is, fixing constants can be moved down from the "meta-level" to the "object level", which allows us to avoid Lifschitz's explicit reference to fixed constants.

We believe that the results in this section are of interest because, apart from showing the expressive power of modified Lifschitz extensions, they allow to establish a relationship between (modified) Lifschitz extensions and the original McCarthy's circumscription, where all the predicate symbols (but the circumscribed one) and all the function symbols are fixed.[19]

To proceed we need the following generalizations of Definitions 25–27.

Definition 47. Let C be a set of function and predicate symbols of **L**. A C-structure s consists of a non-empty domain U_s, an assignment to each n-place predicate symbol P of C of an n-place relation P^s in U_s, and to each n-place function symbol f of C of an n-place function $f^s : U_s^n \to U_s$.[20] If C contains $=$ and $=^s$ is identity, then s is called *normal.*

Definition 48. Let s be a C-structure. We say that an **L**-interpretation w is an s-world, if $U_w = U_s$; for each predicate symbol P of C, $P^w = P^s$; and for each function symbol f of C, $f^w = f^s$. If s is a normal C-structure, then s-worlds are also called normal.

[19] Some embeddings of circumscription into open default theories are presented in the next section.
[20] In particular, if C consists of all the function symbols of **L**, then the class of C-structures coincides with the class of **L**-pre-interpretations, and if C consists of all the function and predicate symbols of **L**, then the class of C-structures coincides with the class of **L**-interpretations.

Definition 49. Let C be a set of function and predicate symbols of **L**. Let s be a C-structure and let (D, A) be a default theory. For any set W of s-worlds let $\Delta^s_{(D,A)}(W)$ be the largest set, V, of s-worlds which are models of A that satisfies the following condition.

For any $\frac{\alpha(x) : M\beta_1(x), \ldots, M\beta_m(x)}{\gamma(x)} \in D$, and any tuple $\boldsymbol{u}$ of elements of U_s, if $\alpha(\boldsymbol{u}) \in \mathrm{Th}_{\mathbf{L}_{U_s}}(V)$, and $\neg\beta_1(\boldsymbol{u}), \ldots, \neg\beta_m(\boldsymbol{u}) \notin \mathrm{Th}_{\mathbf{L}_{U_s}}(W)$, then $\gamma(\boldsymbol{u}) \in \mathrm{Th}_{\mathbf{L}_{U_s}}(V)$.[21]

A set of sentences E is called a (normal) *C-extension* for (D, A) if there exist a (normal) C-structure s and a fixed point W of $\Delta^s_{(D,A)}$ such that $E = \mathrm{Th}_{\mathbf{L}}(W)$. (Note that W must be a set of s-worlds.)

Remark 50. By definition, each normal C-extension is a Lifschitz extension with the *fixed constants* C and vice versa, see [9, Section 6].

Example 51. Let C consist of all the function symbols of **L**. Then each C-extension is an F-extension and vice versa.

Example 52. As we shall see in the proof of Theorem 54 below, each $\{=\}$-extension is a Lifschitz extension and vice versa.

Theorems 53 and 54 below describe a large class of normal C-extensions in terms of modified Lifschitz extensions.

Theorem 53. *Assume that* **L** *contains no function symbols. Then a set of sentences is a normal C-extension for* (D, A) *if and only if for some* U, *it is a modified Lifschitz extension for*

$$\left(D \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\} \cup \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C}, A \cup card_U\right).$$

Theorem 54. *Let* C *contain all the function symbols of* **L**. *Then a set of sentences is a normal C-extension for* (D, A) *if and only if it is a modified Lifschitz extension for*

$$\left(D \cup \left\{\frac{: M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{: MP(\boldsymbol{x})}{P(\boldsymbol{x})}\right\}_{P \in C \cup \{=\}}, A\right).$$

We see that Propositions 43 and 44 follow from Theorems 53 and 54, respectively, with $C = \emptyset$.

Proof of Theorem 53. Let U be a non-empty set. Let F be a universal pre-interpretation and let W be a fixed point of

$$\Delta^F_{(D \cup \{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\}, \{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\}_{P \in C}, A \cup card_U)}.$$

---

[21] Similarly to [9, Proposition 3] it can be shown that this largest class $\Delta^s_{(D,A)}(W)$ always exists.

Substituting $\neg P(x)$ for $\beta(x)$ in Lemma 36, $P \in C \cup \{=\}$, we obtain that the assignment to P is the same relation P^W on U_F in all the elements of W. (Recall that all the elements of W have the same domain U_F.) In particular, $=^W$ is a congruence relation. Since $card_U \in \mathrm{Th}_{\mathbf{L}}(W)$, the cardinality of the set of equivalence classes of $=^W$ is equal to the cardinality of U, if U is finite, and is infinite otherwise. Therefore we may replace U by the set of equivalence classes of $=^W$, which we shall also denote by U. We denote the equivalence class of $u \in U_F$ by $\varepsilon(u)$.

For an F-world $w \in W$, let $\varepsilon(w)$ denote the interpretation over the domain U such that for an n-place predicate symbol P of $\mathbf{L}$, $P^{\varepsilon(w)} = \{(\varepsilon(u_1), \ldots, \varepsilon(u_n)): (u_1, \ldots, u_n) \in P^w\}$. Since the elements of the same equivalence class are indistinguishable in w, $\varepsilon(w)$ is well defined (and is $\mathbf{L}$-equivalent to w). It follows that the restriction of any $\varepsilon(w)$ to the assignments to the elements of C results in the same C-structure s. Moreover, since, by definition, $=^s$ is identity, s is normal.

Let $\varepsilon(W) = \{\varepsilon(w): w \in W\}$. We contend that $\varepsilon(W)$ is a fixed point of $\Delta^s_{(D,A)}$ such that $\mathrm{Th}_{\mathbf{L}}(\varepsilon(W)) = \mathrm{Th}_{\mathbf{L}}(W)$.

Let $w \in W$. It follows from the definition of $\varepsilon(w)$ that for any formula $\phi(x_1, \ldots, x_n)$ of $\mathbf{L}$ and any $u_1, \ldots, u_n \in U_F$, $w \models \phi(u_1, \ldots, u_n)$ if and only if $\varepsilon(w) \models \phi(\varepsilon(u_1), \ldots, \varepsilon(u_n))$. Therefore, $\phi(u_1, \ldots, u_n) \in \mathrm{Th}_{\mathbf{L}_{U_F}}(W)$ if and only if $\phi(\varepsilon(u_1), \ldots, \varepsilon(u_n)) \in \mathrm{Th}_{\mathbf{L}_{U_s}}(\varepsilon(W))$. This, in particular, implies that $\mathrm{Th}_{\mathbf{L}}(\varepsilon(W)) = \mathrm{Th}_{\mathbf{L}}(W)$ (therefore $\varepsilon(W) \models A$), and that the set of the belief normal s-worlds $\varepsilon(W)$ satisfies the condition of Definition 49 for (D, A) and the set of possible normal s-worlds $\varepsilon(W)$. It remains to show that $\varepsilon(W)$ is the largest set of the belief normal s-worlds satisfying that condition.

So, assume to the contrary that there is a set of the belief normal s-worlds V such that $\varepsilon(W)$ is a proper subset of V, and V satisfies the condition of Definition 49 for (D, A) and the set of possible normal s-worlds $\varepsilon(W)$. Since s is a normal C-structure, V satisfies the condition of Definition 49 for

$$\left(D \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\}, \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C}, A \cup card_U\right)$$

and the set of possible worlds $\varepsilon(W)$ as well. For an element $v \in V$, let $\varepsilon^{-1}(v)$ denote the following interpretation based on F. For an n-place predicate symbol P of $\mathbf{L}$, $P^{\varepsilon^{-1}(v)} = \{(u_1, \ldots, u_n): (\varepsilon(u_1), \ldots, \varepsilon(u_n)) \in P^v\}$. Then W is a proper subset of $\varepsilon^{-1}(V) = \{\varepsilon^{-1}(v): v \in V\}$, and the set of the belief F-worlds $\varepsilon^{-1}(V)$ satisfies the condition of Definition 27 for

$$\left(D \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\}, \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C}, A \cup card_U\right)$$

and the set of possible F-worlds W, because, by the definition of ε, for any formula $\phi(x_1, \ldots, x_n)$ of $\mathbf{L}$ and any $u_1, \ldots, u_n \in U_F$, $\phi(u_1, \ldots, u_n) \in \mathrm{Th}_{\mathbf{L}_{U_F}}(\varepsilon^{-1}(V))$ if and only if $\phi(\varepsilon(u_1), \ldots, \varepsilon(u_n)) \in \mathrm{Th}_{\mathbf{L}_{U_s}}(V)$. However, this contradicts the maximality of W.

Conversely, let s be a normal C-structure, and let W be a fixed point of $\Delta^s_{(D,A)}$. We distinguish between the case of finite and infinite U_s. First assume that U_s is infinite. Let F be a universal pre-interpretation with the base U_s. (Since **L** has no function symbols, $F = U_F = U_s$.) Then the belief set W considered as a set of F-worlds satisfies the condition of Definition 27 for

$$\left(D \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\}, \left\{\frac{: M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{: MP(\boldsymbol{x})}{P(\boldsymbol{x})}\right\}_{P \in C}, A \cup card_U\right)$$

and the set of possible F-worlds W. (Recall the elements of W are s-worlds.) Thus the proof in the case of infinite U will be completed if we show that W is the largest set of the belief F-worlds satisfying that condition. Assume to the contrary that there is a set of the belief F-worlds V such that W is a proper subset of V, and V satisfies the condition of Definition 27 for

$$\left(D \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\}, \left\{\frac{: M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{: MP(\boldsymbol{x})}{P(\boldsymbol{x})}\right\}_{P \in C}, A \cup card_U\right)$$

and the set of possible F-worlds W. Let $P \in C \cup \{=\}$. Substituting $\neg P$ for β in Lemma 36, we obtain that V is a set of s-worlds, which contradicts the maximality of W.

Now assume that U_s is finite. Let F be a universal pre-interpretation and let ε be a mapping from U_F onto U_s. For an s-world $w \in W$, let $\varepsilon^{-1}(w)$ denote the interpretation over the domain U_F such that for an n-place predicate symbol P of **L**, $P^{\varepsilon^{-1}(w)} = \{(u_1, \ldots, u_n)\colon (\varepsilon(u_1), \ldots, \varepsilon(u_n)) \in P^w\}$. (In particular, it follows from the definition that $\varepsilon^{-1}(w)$ is **L**-equivalent to w.) Since for a $u \in U_s$, the elements of $\varepsilon^{-1}(u)$ are indistinguishable by $\varepsilon^{-1}(w)$, $\varepsilon^{-1}(w) \models A \cup card_U$. Moreover, since **L** contains no function symbols, $\varepsilon^{-1}(w)$ is an F-world. Let $\varepsilon^{-1}(W) = \{\varepsilon^{-1}(w)\colon w \in W\}$. We intend to show that $\varepsilon^{-1}(W)$ is a fixed point of

$$\left(D \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\}, \left\{\frac{: M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{: MP(\boldsymbol{x})}{P(\boldsymbol{x})}\right\}_{P \in C}, A \cup card_U\right)$$

such that $\mathrm{Th}_{\mathbf{L}}(\varepsilon^{-1}(W)) = \mathrm{Th}_{\mathbf{L}}(W)$.

Obviously, for any formula $\phi(x_1, \ldots, x_n)$ of **L** and any $u_1, \ldots, u_n \in U_F$, $\phi(u_1, \ldots, u_n) \in \mathrm{Th}_{\mathbf{L}_{U_F}}(\varepsilon^{-1}(W))$ if and only if $\phi(\varepsilon(u_1), \ldots, \varepsilon(u_n)) \in \mathrm{Th}_{\mathbf{L}_U}(W)$. Therefore $\mathrm{Th}_{\mathbf{L}}(\varepsilon^{-1}(W)) = \mathrm{Th}_{\mathbf{L}}(W)$. This together with the definition of the assignments of the predicate symbols of $C \cup \{=\}$ in the elements of $\varepsilon^{-1}(W)$ implies that and that the set of the belief F-worlds $\varepsilon^{-1}(W)$ satisfies the condition of Definition 27 for

$$\left(D \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\}, \left\{\frac{: M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{: MP(\boldsymbol{x})}{P(\boldsymbol{x})}\right\}_{P \in C}, A \cup card_U\right)$$

and the set of possible F-worlds $\varepsilon^{-1}(W)$. It remains to show that $\varepsilon^{-1}(W)$ is the largest set of the belief worlds satisfying that condition.

Assume to the contrary that there is a set of the belief worlds V such that

$\varepsilon^{-1}(W)$ is a proper subset of V, and V satisfies the condition of Definition 13 for the default theory

$$\left(D \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\}, \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C}, A \cup card_U\right)$$

and the set of possible F-worlds $\varepsilon^{-1}(W)$. Let $v \in V$. Then for any $P \in C$ and any $u_1, \dots, u_n \in U_F$, $v \models P(u_1, \dots, u_n)$ if and only if $\varepsilon^{-1}(W) \models P(u_1, \dots, u_n)$. Indeed, since W is a set of s-worlds, either $\varepsilon^{-1}(W) \models P(u_1, \dots, u_n)$ or $\varepsilon^{-1}(W) \models \neg P(u_1, \dots, u_n)$. Then the "if" direction follows from Lemma 36 with β being P, and the only if direction follows from Lemma 36 with β being $\neg P$. Let u_1 and u_2 be distinct elements of U_F. We contend that $v \models u_1 = u_2$ if and only if $\varepsilon(u_1) = \varepsilon(u_2)$. Let $\varepsilon(u_1) \neq \varepsilon(u_2)$. Then $\varepsilon^{-1}(W) \models u_1 \neq u_2$, and the "if" direction follows from Lemma 36 with β being $\neq$. Now let $v \models u_1 \neq u_2$. Were $\varepsilon(u_1) = \varepsilon(u_2)$, the number of elements of U_F distinguishable by v would exceed the number of elements of U, in contradiction with $card_U$. This proves the contention.

It follows that $\varepsilon(V)$ is a set of s-worlds which contains W as a proper subset and satisfies the condition of Definition 49 for (D, A) and the set of possible s-worlds W. However, this contradicts the maximality of W. $\square$

Proof of Theorem 54. The proof is similar to that of Theorem 53. Let F be a universal pre-interpretation and let W be a fixed point of

$$\Delta^F_{(D \cup \{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\}_{P \in C \cup \{=\}}, A)} .$$

Let $P \in C \cup \{=\}$. Substituting $\neg P(x)$ for $\beta(x)$ in Lemma 36 we obtain that the assignment to P is the same relation P^W on U_F in all the elements of W. (Recall that all the elements of W have the same domain U_F.) In particular, $=^W$ is a congruence relation. We denote the equivalence class of $u \in U_F$ modulo $=^W$ by $\varepsilon(u)$, and we denote the set of equivalence classes of $=^W$ by $\varepsilon(U_F)$. That is, $\varepsilon(U_F) = \{\varepsilon(u)\}_{u \in U_F}$.

For an F-world $w \in W$, let $\varepsilon(w)$ denote the interpretation over the domain $\varepsilon(U_F)$ such that for an n-place predicate symbol P of $\mathbf{L}$, $P^{\varepsilon(w)} = \{(\varepsilon(u_1), \dots, \varepsilon(u_n)) \colon (u_1, \dots, u_n) \in P^w\}$, and for an n-place function symbol f of $\mathbf{L}$ and for $u_1, \dots, u_n \in U_s$, $f^{\varepsilon(w)}(\varepsilon(u_1), \dots, \varepsilon(u_n)) = \varepsilon(f^w(u_1, \dots, u_n))$. Since the elements of the same equivalence class of $=^W$ are indistinguishable in w, $\varepsilon(w)$ is well defined (and is $\mathbf{L}$-equivalent to w). It follows that the restriction of any $\varepsilon(w)$ to the assignments to the elements of $C \cup \{=\}$ results in the same C-structure s. Moreover, since, by definition, $=^s$ is identity, s is normal. Now exactly as in the proof of Theorem 53 it can be shown that $\varepsilon(W) = \{\varepsilon(w) \colon w \in W\}$ is a fixed point of $\Delta^s_{(D,A)}$ such that $\mathrm{Th}_{\mathbf{L}}(\varepsilon(W)) = \mathrm{Th}_{\mathbf{L}}(W)$:

Let $w \in W$. It follows from the definition of $\varepsilon(w)$ that for any formula $\phi(x_1, \dots, x_n)$ of $\mathbf{L}$ and any $u_1, \dots, u_n \in U_F$, $w \models \phi(u_1, \dots, u_n)$ if and only if $\varepsilon(w) \models \phi(\varepsilon(u_1), \dots, \varepsilon(u_n))$. Therefore, $\phi(u_1, \dots, u_n) \in \mathrm{Th}_{\mathbf{L}_{U_F}}(W)$ if and only if $\phi(\varepsilon(u_1), \dots, \varepsilon(u_n)) \in \mathrm{Th}_{\mathbf{L}_{U_s}}(\varepsilon(W))$. This, in particular, implies that $\mathrm{Th}_{\mathbf{L}}(\varepsilon(W)) =$

$\mathrm{Th}_{\mathbf{L}}(W)$ (therefore $\varepsilon(W) \models A$), and that the set of the belief normal s-worlds $\varepsilon(W)$ satisfies the condition of Definition 49 for (D, A) and the set of possible normal s-worlds $\varepsilon(W)$. It remains to show that $\varepsilon(W)$ is the largest set of the belief normal s-worlds satisfying that condition.

So, assume to the contrary that there is a set of the belief normal s-worlds V such that $\varepsilon(W)$ is a proper subset of V, and V satisfies the condition of Definition 49 for (D, A) and the set of possible normal s-worlds $\varepsilon(W)$. Since s is a normal C-structure, V satisfies the condition of Definition 49 for

$$\left(D \cup \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C \cup \{=\}}, A\right)$$

and the set of possible normal s-worlds $\varepsilon(W)$ as well. For an element $v \in V$, let $\varepsilon^{-1}(v)$ denote the following interpretation based on F. For an n-place predicate symbol P of $\mathbf{L}$, $P^{\varepsilon^{-1}(v)} = \{(u_1, \ldots, u_n)\colon (\varepsilon(u_1), \ldots, \varepsilon(u_n)) \in P^v\}$. Then W is a proper subset of $\varepsilon^{-1}(V) = \{\varepsilon^{-1}(v)\colon v \in V\}$, and the set of the belief F-worlds $\varepsilon^{-1}(V)$ satisfies the condition of Definition 49 for

$$\left(D \cup \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C \cup \{=\}}, A\right)$$

and the set of possible F-worlds W, because, by the definition of ε, for any formula $\phi(x_1, \ldots, x_n)$ of $\mathbf{L}$ and any $u_1, \ldots, u_n \in U_F$, $\phi(u_1, \ldots, u_n) \in \mathrm{Th}_{\mathbf{L}_{U_F}}(V)$ if and only if $\phi(\varepsilon(u_1), \ldots, \varepsilon(u_n)) \in \mathrm{Th}_{\mathbf{L}_{U_s}}(\varepsilon(V))$. However, this contradicts the maximality of W.

Conversely, let s be a normal C-structure, and let W be a fixed point of $\Delta^s_{(D,A)}$. Let F be a universal pre-interpretation such that the cardinality of b_F is not less than the cardinality of U_s. Let ε be a mapping from b_F onto U_s. We extend ε onto U_F by induction as follows. For an n-place function symbol f of $\mathbf{L}$ and $u_1, \ldots, u_n \in U_F$, $\varepsilon(f^F(u_1, \ldots, u_n)) = f^s(\varepsilon(u_1), \ldots, \varepsilon(u_n))$. (Here we use the condition that C contains all the function symbols of $\mathbf{L}$.) For an s-world $w \in W$, let $\varepsilon^{-1}(w)$ denote the interpretation based on F such that for an n-place predicate symbol P of $\mathbf{L}$ and $u_1, \ldots, u_n \in U_F$, $P^{\varepsilon^{-1}(w)} = \{(u_1, \ldots, u_n)\colon (\varepsilon(u_1), \ldots, \varepsilon(u_n)) \in P^w\}$. Let $\varepsilon^{-1}(W) = \{\varepsilon^{-1}(w)\colon w \in W\}$. We intend to show that $\varepsilon^{-1}(W)$ is a fixed point of

$$\left(D \cup \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C \cup \{=\}}, A\right)$$

such that $\mathrm{Th}_{\mathbf{L}}(\varepsilon^{-1}(W)) = \mathrm{Th}_{\mathbf{L}}(W)$.

By definition, $\varepsilon^{-1}(w)$ is an F-world, and a straightforward induction on the complexity of a formula $\phi(x_1, \ldots, x_n)$ shows that for all $u_1, \ldots, u_n \in U_s$, $w \models \phi(u_1, \ldots, u_n)$ if and only if $\varepsilon^{-1}(w) \models \phi(\varepsilon(u_1), \ldots, \varepsilon(u_n))$. In particular, $\varepsilon^{-1}(w) \models A$. Let $P \in C \cup \{=\}$. Since s is a (normal) C-structure, the assignment to P is the same in all the elements of $\varepsilon^{-1}(W)$. Therefore $\mathrm{Th}_{\mathbf{L}}(\varepsilon^{-1}(W)) = \mathrm{Th}_{\mathbf{L}}(W)$, and that the set of the belief F-worlds $\varepsilon^{-1}(W)$ satisfies the condition of Definition 27 for

$$\left(D \cup \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C \cup \{=\}}, A\right)$$

and the set of possible F-worlds $\varepsilon^{-1}(W)$. It remains to show that $\varepsilon^{-1}(W)$ is the largest set of the belief F-worlds satisfying that condition.

Assume to the contrary that there is a set of the belief worlds V such that $\varepsilon^{-1}(W)$ is a proper subset of V, and V satisfies the condition of Definition 27 for

$$\left(D \cup \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C \cup \{=\}}, A\right)$$

and the set of possible F-worlds $\varepsilon^{-1}(W)$. Let $v \in V$, $P \in C \cup \{=\}$, and let $u_1, \ldots, u_n \in U_F$. We contend that $v \models P(u_1, \ldots, u_n)$ if and only if $\varepsilon^{-1}(W) \models P(u_1, \ldots, u_n)$. As we have seen earlier, $\varepsilon^{-1}(W) \models P(u_1, \ldots, u_n)$ if and only if $W \models P(\varepsilon(u_1), \ldots, \varepsilon(u_n))$, and, since all the elements of W are (normal) s-worlds, either $W \models P(\varepsilon(u_1), \ldots, \varepsilon(u_n))$ or $W \models \neg P(\varepsilon(u_1), \ldots, \varepsilon(u_n))$. Therefore the "if" direction of the contention follows from Lemma 36 with β being P, and the only if direction follows from Lemma 36 with β being $\neg P$. Thus, $\varepsilon(v)$ is an s-structure.

It follows that $\varepsilon(V)$ is the set of s-worlds which contains W as a proper subset. Also the set of the belief s-worlds $\varepsilon(V)$ satisfies the condition of Definition 49 for (D, A) and the set of possible s-worlds W, which contradicts the maximality of W. □

Next we describe normal C-extensions in terms of Lifschitz extensions.

Theorem 55. *A set of sentences is a normal C-extension for (D, A) if and only if it is a Lifschitz extension for*

$$\left(D \cup \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C} \cup \left\{\frac{: Mf(x) \neq x}{f(x) \neq x}, \frac{: Mf(x) = x}{f(x) = x}\right\}_{f \in C}, A\right).$$

For the proof of Theorem 4 we need to extend Lemma 36 to arbitrary C-structures.

Lemma 56. *Let (D, A) be a default theory and let D contain a normal default without prerequisites $\frac{: M\beta(x)}{\beta(x)}$. Let s be a C-structure and let W be a set of s-worlds. If for some $w \in W$ and some tuple $\boldsymbol{u}$ of elements of U_s, $w \models \beta(\boldsymbol{u})$, then for any $v \in \Delta^s_{(D,A)}(W)$, $v \models \beta(\boldsymbol{u})$.*

Proof. The proof follows immediately from the definition of $\Delta^s_{(D,A)}$ (Definition 49). □

Proof of Theorem 55. The proof is similar to those of Theorems 53 and 54. Let U be a non-empty set and let W be a fixed point of

$$\Delta^U_{(D \cup \{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\}_{P \in C} \cup \{\frac{: Mf(x) \neq x}{f(x) \neq x}, \frac{: Mf(x) = x}{f(x) = x}\}_{f \in C}, A)}\,.$$

For any $P \in C$ ($f \in C$), substituting $\neg P(x)$ ($f(x) \neq x$) for $\beta(x)$ in Lemma 56, we obtain that the assignment to P (f) is the same relation P^W (function f^W) in all the elements of W. (Recall that all the elements of W have the same domain U.) Therefore, the restriction of interpretations of W to the symbols of C results in the same normal C-structure s ($U_s = U$). That is, W is a set of normal s-worlds.

Since W is a fixed point of

$$\Delta^U_{(D \cup \{\frac{:M\neg P(x)}{\neg P(x)}, \frac{:MP(x)}{P(x)}\}_{P \in C} \cup \{\frac{:Mf(x) \neq x}{f(x) \neq x}, \frac{:Mf(x) = x}{f(x) = x}\}_{f \in C}, A)},$$

the set of the belief s-worlds W satisfies the condition of Definition 49 for (D, A) and the set of possible normal s-worlds W. It remains to show that W is the largest set of the belief worlds satisfying that condition.

Assume to the contrary that there is a set of the belief worlds V such that W is a proper subset of V, and V satisfies the condition of Definition 49 for (D, A) and the set of possible normal s-worlds W. By definition, for any $\frac{\alpha(x) : M\beta_1(x), \ldots, M\beta_m(x)}{\gamma(x)} \in D$, and any tuple $\boldsymbol{u}$ of elements of U ($= U_s$), if $\alpha(\boldsymbol{u}) \in \mathrm{Th}_{\mathrm{L}_U}(V)$, and $\neg\beta_1(\boldsymbol{u}), \ldots, \neg\beta_m(\boldsymbol{u}) \notin \mathrm{Th}_{\mathrm{L}_U}(W)$, then $\gamma(\boldsymbol{u}) \in \mathrm{Th}_{\mathrm{L}_U}(V)$. Also, since the assignments to the predicate and function symbols belonging to C are the same in all the elements of V, for any $P \in C$ and any tuple $\boldsymbol{u}$ of elements of U, if $P(\boldsymbol{u}) \notin \mathrm{Th}_{\mathrm{L}_U}(W)$, then $\neg P(\boldsymbol{u}) \in \mathrm{Th}_{\mathrm{L}_U}(V)$, and $\neg P(\boldsymbol{U}) \notin \mathrm{Th}_{\mathrm{L}_U}(W)$, then $P(\boldsymbol{u}) \in \mathrm{Th}_{\mathrm{L}_U}(V)$. Similarly, for any $f \in C$, any tuple $\boldsymbol{u}$ of elements of U and any $u \in U$, if $f(\boldsymbol{u}) = u \notin \mathrm{Th}_{\mathrm{L}_U}(W)$, then $f(\boldsymbol{u}) \neq u \in \mathrm{Th}_{\mathrm{L}_U}(V)$, and if $f(\boldsymbol{u}) \neq u \notin \mathrm{Th}_{\mathrm{L}_U}(W)$, then $f(\boldsymbol{u}) = u \in \mathrm{Th}_{\mathrm{L}_U}(V)$. It follows that the set of the belief U-worlds V satisfies the condition of Definition 20 for (D, A) and the set of possible U-worlds W, which contradicts the maximality of W.

Conversely, let s be a normal C-structure, and let W be a fixed point of $\Delta^s_{(D,A)}$. Let $U = U_s$. As we saw in the proof of the "only if" part of the theorem, the set of the belief U-worlds W satisfies the condition of Definition 20 for

$$\left(D \cup \left\{\frac{:M\neg P(x)}{\neg P(x)}, \frac{:MP(x)}{P(x)}\right\}_{P \in C} \cup \left\{\frac{:Mf(x) \neq x}{f(x) \neq x}, \frac{:Mf(x) = x}{f(x) = x}\right\}_{f \in C}, A\right)$$

and the set of possible U-worlds W. The proof will be completed, if we show that W is the largest set of belief U-worlds satisfying that condition.

Assume to the contrary that there is a set of the belief U-worlds V such that W is a proper subset of V, and V satisfies the condition of Definition 20 for the default theory

$$\left(D \cup \left\{\frac{:M\neg P(x)}{\neg P(x)}, \frac{:MP(x)}{P(x)}\right\}_{P \in C} \cup \left\{\frac{:Mf(x) \neq x}{f(x) \neq x}, \frac{:Mf(x) = x}{f(x) = x}\right\}_{f \in C}, A\right)$$

and the set of possible U-worlds W. Then for any $P \in C$, any $v \in V$ and any tuple $\boldsymbol{u}$ of elements of U, $v \models P(\boldsymbol{u})$ if and only $W \models P(\boldsymbol{u})$. That is, V is a set of s-worlds. (Recall that interpretations of W are s-worlds. Therefore $W \models P(\boldsymbol{u})$ if and only if $w \models P(\boldsymbol{u})$ for some $w \in W$.) Indeed, the "if" direction follows from Lemma 56 with β being P, and the only if direction follows from Lemma 56 with β being $\neg P$.

Similarly, for any $f \in C$, any $v \in V$, any tuple $\boldsymbol{u}$ of elements of U and any $u \in U$, $v \models f(\boldsymbol{u}) = u$ if and only $W \models f(\boldsymbol{u}) = u$. By definition, for any $\frac{\alpha(x) : M\beta_1(x), \ldots, M\beta_m(x)}{\gamma(x)} \in D$, and any tuple $\boldsymbol{u}$ of elements of U, if $\alpha(\boldsymbol{u}) \in \mathrm{Th}_{\mathbf{L}_U}(V)$, and $\neg\beta_1(\boldsymbol{u}), \ldots, \neg\beta_m(\boldsymbol{u}) \notin \mathrm{Th}_{\mathbf{L}_U}(W)$, then $\gamma(\boldsymbol{u}) \in \mathrm{Th}_{\mathbf{L}_U}(V)$. That is, the set of the belief normal s-worlds V satisfies the condition of Definition 49 for (D, A) and the set of possible normal s-worlds W, which contradicts the maximality of W. □

Now, in view of Theorems 53 and 54 one can try to express normal C-extensions in terms of modified Lifschitz extensions. However this cannot be done straightforward, because modified Lifschitz extensions are based on interpretations where assignments to function symbols are fixed in a very special independent manner. One possible way to avoid such an independence is first to allow all the functions to vary by passing to the defining expansion of $\mathbf{L}$, and then to fix the defining predicates of the functions belonging to C.

Proposition 57. *Let C be a set of function and predicate symbols of $\mathbf{L}$. Then a set of sentences E is a normal C-extension for a default theory (D, A) if and only if for some U, E^{P} is a modified Lifschitz extension for*

$$\left(D^{\mathrm{P}} \cup \left\{\frac{: Mx_1 \neq x_2}{x_1 \neq x_2}\right\} \cup \left\{\frac{: M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{: MP(\boldsymbol{x})}{P(\boldsymbol{x})}\right\}_{P \in C} \cup \left\{\frac{: M\neg P_f(\boldsymbol{x})}{\neg P_f(\boldsymbol{x})}, \frac{: MP_f(\boldsymbol{x})}{P_f(\boldsymbol{x})}\right\}_{f \in C}, A^{\mathrm{D}} \cup card_U\right).$$

Proof. Let $C^{\mathrm{P}} = (C - \{f\}_{f \in C}) \cup \{P_f\}_{f \in C}$. Exactly as in the proof of Proposition 24 it can be shown that E is a normal C-extension for (D, A) if and only if E^{P} is a normal C^{P}-extension for $(D^{\mathrm{P}}, A^{\mathrm{D}})$. By Theorem 55 and the definition of C^{P}, E^{P} is a normal C^{P}-extension for $(D^{\mathrm{P}}, A^{\mathrm{D}})$ if and only if it is a Lifschitz extension for

$$\left(D^{\mathrm{P}} \cup \left\{\frac{: M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{: MP(\boldsymbol{x})}{P(\boldsymbol{x})}\right\}_{P \in C} \cup \left\{\frac{: M\neg P_f(\boldsymbol{x})}{\neg P_f(\boldsymbol{x})}, \frac{: MP_f(\boldsymbol{x})}{P_f(\boldsymbol{x})}\right\}_{f \in C}, A^{\mathrm{D}}\right).$$

Now the proof follows immediately from Theorem 53. □

Proposition 58. *Let C be a set of function and predicate symbols of $\mathbf{L}$. Then a set of sentences E is a normal C-extension for a default theory (D, A) if and only if E^{P} is a modified Lifschitz extension for*

$$\left(D^{\mathrm{P}} \cup \left\{\frac{: M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{: MP(\boldsymbol{x})}{P(\boldsymbol{x})}\right\}_{P \in C \cup \{=\}} \cup \left\{\frac{: M\neg P_f(\boldsymbol{x})}{\neg P_f(\boldsymbol{x})}, \frac{: MP_f(\boldsymbol{x})}{P_f(\boldsymbol{x})}\right\}_{f \in C}, A^{\mathrm{D}}\right).$$

The proof of Proposition 58 differs from that of Proposition 57 only in replacing the words "Theorem 53" by "Theorem 54". We leave it to the reader.

Proposition 60 below shows that normal C-extensions can be expressed in terms of modified Lifschitz extensions in a more direct manner. Namely, instead of passing to the defining expansion of $\mathbf{L}$ and fixing the defining predicates of the functions belonging to C, it suffices to "define" the varied function symbols, only. We shall need the following generalization of Definition 23.

Definition 59. Let C be a set of function and predicate symbols of $\mathbf{L}$. Let P be a mapping from the set of all function symbols f of $\mathbf{L}$ which do not belong to C to a list of new predicate symbols P_f. We assume that P is one-to-one and if f is an n-place function symbol, then P_f is an $n+1$-place (defining) predicate symbol. We call the language $\mathbf{L}_{\mathrm{D}_C}$ obtained from $\mathbf{L}$ by deleting its function symbols not belonging to C and extending its predicate symbols with $\{P_f\colon f \not\in C\}$ the (C-) *defining expansion* of $\mathbf{L}$.[22] For $f \not\in C$, the sentence $\forall x_1 \cdots \forall x_n \exists ! x P_f(x_1, \ldots, x_n, x)$ is called a (C-) *defining axiom*. The set of all C-defining axioms is denoted by $\mathbf{D}_C$.

For a formula ϕ of $\mathbf{L}$ we define its translation, denoted ϕ^{P_C}, into $\mathbf{L}_{\mathrm{D}_C}$, by induction, as follows. Consider a sequence $\phi_0, \phi_1, \ldots$ of formulas over the language extended with the new predicate symbols, such that ϕ_0 is ϕ, and ϕ_{i+1} results from ϕ_i in the following manner. If ϕ_i does not contain function symbols not belonging to C, then the sequence terminates at ϕ_i. Otherwise, let $f(t_1, \ldots, t_n)$ be the leftmost term in ϕ_i such that $f \not\in C$ and all t_i are variables; and let $P(\ldots, f(t_1, \ldots, t_n), \ldots)$ be the atomic subformula of ϕ_i that contains that term. Then ϕ_{i+1} is obtained from ϕ_i by replacing $P(\ldots, f(t_1, \ldots, t_n), \ldots)$ with $\exists x(P_f(t_1, \ldots, t_n, x) \wedge P(\ldots, x, \ldots))$. Since the number of function symbols of ϕ_{i+1} which do not belong to C is one less than the number of those in ϕ_i, the above sequence must terminate. The last formula in the sequence is ϕ^{P_C}. For a set of $\mathbf{L}$-formulas X we define a set of $\mathbf{L}_{\mathrm{D}_C}$-formulas X^{P_C} by $X^{\mathrm{P}_C} = \{\phi^{\mathrm{P}_C}\colon \phi \in X\}$ and for a set D of defaults over $\mathbf{L}$ we define a set D^{P_C} of defaults over $\mathbf{L}_{\mathrm{D}_C}$ by

$$D^{\mathrm{P}_C} = \left\{ \frac{\alpha^{\mathrm{P}_C}(x) : M\beta_1^{\mathrm{P}_C}(x), \ldots, M\beta_m^{\mathrm{P}_C}(x)}{\gamma^{\mathrm{P}_C}(x)} : \frac{\alpha(x) : M\beta_1(x), \ldots, M\beta_m(x)}{\gamma(x)} \in D \right\}.$$

Finally, for the first-order theory X over $\mathbf{L}$, the (C-) *defining expansion* of X is a theory X^{D_C} over $\mathbf{L}_{\mathrm{D}_C}$ defined by $X^{\mathrm{D}_C} = X^{\mathrm{P}_C} \cup \mathbf{D}_C$.

Proposition 60. *Let C be a set of function and predicate symbols of $\mathbf{L}$ and let (D, A) be a default theory. Then a set of sentences E is a normal C-extension for (D, A) if and only if E^{P_C} is a modified Lifschitz extension for*

$$\left(D^{\mathrm{P}_C} \cup \left\{ \frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)} \right\}_{P \in C \cup \{=\}}, A^{\mathrm{D}_C} \right).$$

Proof. The proof is similar to those of Propositions 57 and 58. Exactly as in the proof of Proposition 24 it can be shown that E is a normal C-extension for (D, A) if and only if E^{P_C} is a normal C-extension for $(D^{\mathrm{P}_C}, A^{\mathrm{D}_C})$. Since C contains all the function symbols of $\mathbf{L}_{\mathrm{D}_C}$, by Theorem 54, E^{P_C} is a normal C-extension for $(D^{\mathrm{P}_C}, A^{\mathrm{D}_C})$ if and only if it is a modified Lifschitz extension for

[22] In particular, if C contains no function symbols, then $\mathbf{L}_{\mathrm{D}_C} = \mathbf{L}_{\mathrm{D}}$, see Definition 23.

$$\left(D^{\mathrm{P}_C} \cup \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{: MP(x)}{P(x)}\right\}_{P \in C \cup \{=\}}, A^{\mathrm{D}_C}\right). \quad \square$$

10. Modified Lifschitz extensions and circumscription

It is known from [9] that circumscription can be expressed in terms of normal extensions with fixed constants. Since, by Propositions 57, 58, and 60, normal extensions with fixed constants can be expressed in terms of modified Lifschitz extensions, circumscription can be expressed in terms of modified Lifschitz extensions as well. For the proofs of the expressibility results in this section we need to extend Definition 35 and Lemma 37 to arbitrary C-structures.

Definition 61. Let D be a set of normal defaults without prerequisites. Let C be a set of function and predicate symbols and let s be a C-structure. We say that an s-world w is *D-maximal* if there is no s-world w' such that for all $\frac{: M\beta(x)}{\beta(x)} \in D$, $\beta^w \subseteq \beta^{w'}$; and for some $\frac{: M\beta(x)}{\beta(x)} \in D$, β^w is a proper subset of $\beta^{w'}$. We say that s-worlds w_1 and w_2 are *β-equivalent*, denoted $w_1 \sim_{\mathrm{D}} w_2$, if $\beta^{w_1} = \beta^{w_2}$, for all $\frac{: M\beta(x)}{\beta(x)} \in D$. (It immediately follows from the definition that $\sim_D$ is an equivalence relation.)

The proof of the following lemma is similar to that of Lemma 37 and will be omitted.

Lemma 62. *Let C be a set of function and predicate symbols and let s be a C-structure. Let D be a set of normal defaults without prerequisites. A set of s-worlds is a fixed point of $\Delta^s_{(D,A)}$ if and only if it is an equivalence class of the restriction of $\sim_D$ to the D-maximal s-worlds which are models of A.*

Propositions 63–65 below are different translations of [9, Proposition 7] into modified Lifschitz extensions.

Proposition 63. *Let $\boldsymbol{P}$ be a set of predicate symbols, and let C be a set of function and predicate symbols disjoint with $\boldsymbol{P}$. A sentence ϕ is entailed by the circumscription of $\boldsymbol{P}$ in $A(\boldsymbol{P})$ with the fixed constants C if and only if for all U, ϕ^{P} belongs to all modified Lifschitz extensions for*

$$\left(\left\{\frac{: M\neg P(x)}{\neg P(x)}\right\}_{P \in \boldsymbol{P} \cup \{=\}} \cup \left\{\frac{: M\neg P(x)}{\neg P(x)}, \frac{:MP(x)}{P(x)}\right\}_{P \in C} \cup \left\{\frac{: M\neg P_f(x)}{\neg P_f(x)}, \frac{: MP_f(x)}{P_f(x)}\right\}_{f \in C}, A^{\mathrm{D}} \cup card_U\right).$$

Proposition 64. *Let $\boldsymbol{P}$ be a set of predicate symbols, and let C be a set of function and predicate symbols disjoint with $\boldsymbol{P}$. A sentence ϕ is entailed by the circumscrip-*

*tion of **P** in A(**P**) with the fixed constants C if and only if ϕ^{P} belongs to all modified Lifschitz extensions for*

$$\left(\left\{\frac{:M\neg P(x)}{\neg P(x)}\right\}_{P\in \boldsymbol{P}} \cup \left\{\frac{:M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{:MP(\boldsymbol{x})}{P(\boldsymbol{x})}\right\}_{P\in C\cup\{=\}} \cup \left\{\frac{:M\neg P_f(\boldsymbol{x})}{\neg P_f(\boldsymbol{x})}, \frac{:MP_f(\boldsymbol{x})}{P_f(\boldsymbol{x})}\right\}_{f\in C}, A^{\mathrm{D}}\right).$$

Proposition 65. *Let **P** be a set of predicate symbols, and let C be a set of function and predicate symbols disjoint with **P**. A sentence ϕ is entailed by the circumscription of **P** in A(**P**) with the fixed constants C if and only if ϕ^{P_C} belongs to all modified Lifschitz extensions for*

$$\left(\left\{\frac{:M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}\right\}_{P\in \boldsymbol{P}} \cup \left\{\frac{:M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{:MP(\boldsymbol{x})}{P(\boldsymbol{x})}\right\}_{P\in C\cup\{=\}}, A^{\mathrm{D}_C}\right).$$

Proof of Proposition 63. The proof is similar to that of [9, Proposition 7]. By [8, Proposition 1], a sentence ϕ is entailed by the circumscription of **P** in $A(\boldsymbol{P})$ with the fixed constants C if and only if it is satisfied by all **P**-minimal C-structures which are normal models of $A(\boldsymbol{P})$. Obviously, a model of $A(\boldsymbol{P})$ is **P**-minimal if and only if it is $\{\frac{:M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}\}_{P\in\boldsymbol{P}}$-maximal. By Lemma 62, the class of such models coincides with the union of all (normal) fixed points of all

$$\Delta^{s}_{(\{\frac{:M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}\}_{P\in\boldsymbol{P}},\{A(\boldsymbol{P})\})},$$

where s is a C-structure, and the result follows from Proposition 57. □

The proof of Propositions 64 and 65 differs from that of Proposition 63 only in replacing the words "Proposition 57" by "Proposition 58" and "Proposition 60", respectively. We leave them to the reader.

Now the original McCarthy's circumscription of one predicate can be very nicely expressed in terms of modified Lifschitz extensions.

Proposition 66. *A sentence ϕ is entailed by the circumscription of predicate P in A(P) if and only if ϕ belongs to all modified Lifschitz extensions for*

$$\left(\left\{\frac{:M\neg P(x)}{\neg P(x)}\right\} \cup \left\{\frac{:M\neg Q(\boldsymbol{x})}{\neg Q(\boldsymbol{x})}, \frac{:MQ(\boldsymbol{x})}{Q(\boldsymbol{x})}\right\}_{Q\neq P}, A\right).$$

Proof. The proof follows immediately from Proposition 65, because if C contains all the function and predicate symbols, but P, then A^{D_C} is A itself. □

In the above propositions, dealing with normal interpretations, we implicitly used the assumption that = does not belong to **P**. Indeed, by definition, = is fixed in normal interpretations (which are $\neq$-maximal). One of the features of our approach is that it allows = to vary. Thus, we can try to circumscribe equality, by substituting = for P in Proposition 66. Note that for this we have first to pass to the defining expansion, because we cannot use equality anymore for fixing function symbols. Equivalently, instead of passing to the defining expansion of **L**, we may assume that **L** does not have function symbols.

Let $EQ(=)$ be the set of all the equality axioms. The "syntactical" circumscription of = in $EQ(=)$ states that = is the minimal congruence relation, which also follows from the definition of equality without applying circumscription.[23] If in Proposition 66 we substitute = for P, we obtain that *a sentence ϕ is entailed by the circumscription of = if and only if ϕ belongs to all modified Lifschitz extensions for*

$$\left(\left\{ \frac{: Mx_1 \neq x_2}{x_1 \neq x_2} \right\} \cup \left\{ \frac{: M\neg P(\boldsymbol{x})}{\neg P(\boldsymbol{x})}, \frac{: MP(\boldsymbol{x})}{P(\boldsymbol{x})} \right\}_{``P \neq = ''}, EQ(=) \right),$$

or, equivalently, that ϕ is satisfied by all normal interpretations. That is, semantically, the circumscribed equality is identity, which perfectly matches the intuition. Since the syntactical counterpart of satisfiability by all the normal interpretations is provability in the first-order predicate calculus, as we already know, the result of circumscription of = in $EQ(=)$ is = itself.

11. Concluding remarks

In this paper we pointed out some obvious lacks of definitions of extensions for open default theories known from the literature, and considered a possible definition which is less vulnerable. This definition, basically, follows Lifschitz's semantical approach [9], whereas other definitions of extensions for open default theories known from the literature are syntactical. In particular, whereas in Reiter's and Poole's approaches the free variables of a default are treated as meta-variables for closed terms, in Lifschitz's and our approaches they are treated as the theory object variables. The reason for choosing semantical definition of extensions is that it provides a complete (but indirect) description of the theory objects, which in turn, implies the domain closure assumption.

The difference between ours and Lifschitz's original definition is the replacement of sets of normal interpretations over the same possibly finite domain by sets of interpretations with the same universal pre-interpretation. Whereas normal and universal non-normal interpretation approaches are equivalent in the case of closed default theories considered by Guerreiro and Casanova in [4], the restriction to normal interpretations implies undesirable consequences in the case of open default theories. In general, the restriction to normal interpretations seems unnatural, because, for example, equality is definable, if the underlying language has only finitely many predicate and function symbols.

The second difference between Lifschitz's and our approaches is that Lifschitz's definition, requiring the assignment of identity to the equality relation, empha-

[23] The fact that circumscribing equality does not produce anything new was pointed out in [3].

sizes the predicate structure of the interpretation, whereas our approach is based on the compatibility of the assignments to function symbols. An indirect support for the compatibility requirement is that the assignment to functions in the worlds of Kripke models for the first-order modal logics must be compatible (and can be assumed universal), and there is a tight relationship between closed default theories and nonmonotonic modal logics, see [6].

Finally we would like to note that the evidence for the rightness of our approach is only intuitive. There is no (and cannot be any) conclusive formal criteria for the rightness of a definition of extensions for open default theories. The question "which definition is the right one?" can be answered only by a field test.

Acknowledgement

I am deeply indebted to G. Schwarz for his careful reading of the last versions of the paper, for his help in fixing some incorrect statements, and many extremely helpful discussions which led to the present form of the paper. Also I would like to thank the anonymous referees for their criticism and suggestions, M. Tiomkin for pointing out some flaws in an earlier version of the paper, and N. Francez for his comments and for calling my attention to reference [10].

References

[1] C.C. Chang and H.J. Keisler, *Model Theory* (North Holland, Amsterdam, 1990).

[2] D.W. Etherington, Relating default reasoning and circumscription, in: *Proceedings IJCAI-87*, Milan, Italy (Morgan Kaufmann, San Mateo, CA, 1987) 489–494.

[3] D.W. Etherington, R.E. Mercer and R. Reiter, On the adequacy of predicate circumscription for closed-world reasoning, *Comput. Intell.* **1** (1985) 11–15.

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

[5] G.E. Hughes and M.J. Cresswell, *An Introduction to Modal Logic* (Methuen, London, 1972).

[6] M. Kaminski, Interpretations of open default theories in non-monotonic logics, *Annals Math. Artif. Intell.* **9** (1993) 307–325.

[7] K. Konolige, Quantification in autoepistemic logic, *Fund. Inform.* **XV** (1991) 27–300.

[8] V. Lifschitz, Computing circumscription, in: *Proceedings IJCAI-87*, Milan, Italy (Morgan Kaufmann, San Mateo, CA, 1987) 121–127.

[9] V. Lifschitz, On open defaults, in: J.W. Lloyd, ed., *Computational Logic: Symposium Proceedings* (Springer-Verlag, Berlin, 1990) 80–9.

[10] J.W. Lloyd, *Foundations of Logic Programming, Second Extended Edition* (Springer-Verlag, Berlin, 1993).

[11] J. McCarthy, Circumscription—a form of non-monotonic reasoning, *Artif. Intell.* **13** (1980) 27–39.

[12] E. Mendelson, *Introduction to Mathematical Logic* (Wadsworth and Brooks, Monterey, CA, 1987).

[13] J.D. Monk, *Mathematical Logic* (Springer-Verlag, New York, 1976).

[14] D. Poole, A logical framework for default reasoning, *Artif. Intell.* **36** (1988) 27–47.
[15] R. Reiter, A logic for default reasoning, *Artif. Intell.* **13** (1980) 81–132.
[16] J.A. Robinson, A machine oriented logic based on the resolution principle, *J. ACM* **12** (1965) 25–41.