ELSEVIER

Artificial Intelligence 93 (1997) 169-199

Artificial Intelligence

Definability and commonsense reasoning

Gianni Amati [a,*], Luigia Carlucci Aiello [b,1], Fiora Pirri [b,2]

[a] *Fondazione Ugo Bordoni, via Baldassarre Castiglione 59, 00142 Roma, Italy*

[b] *Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", via Salaria 113, 00198 Roma, Italy*

Received March 1996; revised September 1996

Abstract

The definition of concepts is a central problem in commonsense reasoning. Many themes in nonmonotonic reasoning concern implicit and explicit definability. Implicit definability in nonmonotonic logic is always relative to the context—the current theory of the world. We show that fixed point equations provide a generalization of explicit definability, which correctly captures the relativized context. Theories expressed within this logical framework provide implicit definitions of concepts. Moreover, it is possible to derive these fixed points entirely within the logic. © 1997 Elsevier Science B.V.

Keywords: Commonsense reasoning; Definability; Fixed points; Logic of provability; Default logic; Contextual reasoning; Self-reference

1. Introduction and motivations

Concepts play a central role in commonsense reasoning. The classical view consists of postulating a definition of a concept, in terms of necessary and sufficient conditions, in which the properties used as *definientes* are independent of the *definiendum*.

Positivists pointed out that if a concept cannot be defined via necessary and sufficient conditions it is not a scientific concept. It turns out, however that there are concepts, like *game*—as noted by Wittgenstein—that do not seem to have a common core which could be characterized by a set of necessary conditions (see [15] for a discussion).

* Corresponding author. E-mail: gba@fub.it. Work carried out in the framework of the agreement between the Italian PT Administration and the Fondazione Ugo Bordoni.

[1] E-mail: aiello@dis.uniroma1.it.

[2] E-mail: pirri@dis.uniroma1.it.

PII S0004-3702(96)00049-5

On the other hand, concepts involving natural kinds like *bird*, *lemon*, etc., possess necessary but not sufficient conditions. This point is stressed in *prototype theory*, where Eleanor Rosch is one of the main proponents (see e.g. [32]). In connection with this view on "natural kinds", Reiter [41] suggests resorting to sufficient conditions postulated via some linguistic pattern like "normally", "typically", or "assume by default", which appeals to the context to which the definition is relativized. And, in fact, these linguistic patterns mirror the role of defaults; namely, a default theory $\langle W, D\rangle$ accounts for a set of classical necessary conditions W and a set of "sufficient" default conditions D.

Likewise, in almost all cases reported in the literature (e.g. [22,24,29]), solving a circumscription axiomatization (by finding an equivalent first-order theory) yields definitions for the predicate being minimized. The circumscription axiom acts like an implicit sufficient condition (just as, for Reiter, the default rules act like sufficient conditions) whereby the theory is implicitly defining the concept. Solving the circumscription amounts to finding an explicit definition for the predicates being minimized.

The notions of implicit and explicit definition of a predicate, with respect to a theory, are formalized in first-order logic through Beth's definability theorem (see e.g. [6]), which shows that the two notions are equivalent. Moreover, from an implicit definition of a predicate an explicit definition can be found by constructing an interpolant formula. An immediate consequence of Beth's theorem is that implicit definitions in first-order logic state the uniqueness of the predicate being defined. Therefore, for example, if we state that $S \vdash \forall x\ (\text{RAVEN}(x) \equiv \text{BLACK}(x))$, then the only black things we can talk about in S are ravens. Whenever the context changes, for example by adding new axioms, then by the monotonicity of first-order logic, one cannot compatibly update the definition of the concept. Tarski discusses the connection between the notions concerned with definability and those concerned with deduction in [54].

On the other hand, only "well-behaved" predicates are implicitly definable; in general we may lack either sufficient or necessary conditions to define a concept; in such cases no explicit definition can be drawn. McCarthy's circumscription [29] circumvents this strong behaviour of definability in first-order logic because the circumscription axiom is weaker than implicit definability (see [11] for a discussion). As noted above, since the explicit definition is achieved by minimization, uniqueness is no longer a strong constraint because it is relativized to the theory on which circumscription is applied. In other words, there is no total commitment to the given definition.

Consider, for example, the following theory T.

$$\begin{aligned} &\text{ONTABLE}(a) \vee \text{ONTABLE}(b), \\ &\forall x\ \text{RED}(x) \rightarrow \text{ONTABLE}(x). \end{aligned} \tag{1}$$

The circumscription of ONTABLE in T yields a first-order formula because ONTABLE is separable (see [21]), that is

$$\begin{aligned} &(\forall x\ \text{ONTABLE}(x) \equiv (\text{RED}(x) \vee x = a) \wedge Z) \vee \\ &(\forall x\ \text{ONTABLE}(x) \equiv (\text{RED}(x) \vee x = b) \wedge Z') \end{aligned} \tag{2}$$

where Z and Z' are formulae not containing ONTABLE (see [21, Theorem 1]).

We note that in such a case we do not get an implicit definition of ONTABLE, as there are at least two minimal nonisomorphic models for it. Therefore we do not get an explicit definition in Beth sense. Despite the lack of a classical explicit definition, we have obtained a *disjunction of two definitions* for the predicate ONTABLE. The example clearly shows that circumscription is weaker than Beth's implicit definability and we are no longer committed to uniqueness. A first-order disjunction of definitions is a nice generalization of definability, though it does not tell how a single explicit definition can be obtained from weak implicit conditions. Lin [23] shows that the circumscription of a theory, axiomatizing the effects of indeterminate actions, may yield a very large disjunction of successor state axioms. To overcome this problem, Lin proposes a transformation that breaks the disjunction and yields different successor state axioms. The interesting contribution is that his transformation is performed by introducing a suitable predicate which is used, in a sense, to name the contexts in which the different effects of the performed action are realized.

As another example of how implicit and explicit definability enters into nonmonotonic reasoning, observe that an approach to logic programming semantics for the nonmonotonic negation as failure operator is the Clark completion, which treats a logic program as a set of necessary conditions and "completes" this program by adding suitable sufficient conditions yielding explicit definitions for the program predicates. Reiter [40] shows how, in some cases, the Clark completion is the result of circumscribing the program.

The upshot of the foregoing discussion is that many themes in nonmonotonic reasoning concern implicit and explicit definability. In fact there are very good computational reasons for wanting explicit definitions, because then we can do efficient theorem proving by substitution of the definientes for the definiendum in any theorem to be proved. However, while it seems clear how to weaken implicit definability in nonmonotonic reasoning (via defaults or circumscription) so far there is no solution to the problem of how to get, in general, an explicit definition from the implicit sufficient conditions stated through a nonmonotonic theory. The main contribution of this paper is to show one way to solve this problem.

Returning to the above theory (1), in order to obtain one of the disjuncts disregarding the others, we should be able to express that, relative to a context C, the necessary and sufficient conditions for an object to be on the table are that either it is red or it is the object a. Analogously there is a context C' where the object is b. The role played by contexts should be the following: each definition in the disjunction should refer to a different context.

To subsume a context of reasoning (or a context of discourse) in a theory axiomatizing a certain state of affairs, we need a language that resorts to a kind of self-referential ability. That is, we need to formalize, in the language, expressions that can be reasoned about in that same language itself. A statement in which the context can be *explicitly* taken into account is a *commonsense statement*. And, in fact, commonsense reasoning naturally relies on a current state of affairs: the minimum resource of information at hand.

This form of relativization to the context is often carried out by binding commonsense statements to the belief set of an agent. So the self-referential ability is lifted from the

commonsense language to one in which two distinct (or even more) levels of reasoning are formulated: the one where the agent draws conclusions from the initial assumptions and the one where the conclusions drawn are compared with the context, which, in general, is a metalogical structure, i.e., a computational object external to the language in which the agent is reasoning. The following definition is, in this sense, paradigmatic:

$$\Gamma \equiv Cn_{\Lambda}(I \cup \{\Diamond\alpha \mid \neg\alpha \notin \Gamma\}) \tag{3}$$

which says that if I is a nonmonotonic theory then the nonmonotonic consequences of I are obtained by taking the deductive closure—in the logic Λ—of I, together with all the formulae consistent with Γ, where $\Diamond\alpha$ is, in fact, interpreted as "it is consistent to assume α in Γ". Although Γ is a context, in the sense that it is considered to be the belief set of an agent, the above expression does not belong to the agent's language. When the above schema is used to characterize default reasoning, we have three levels of discourse. One for the default theory, one for the translation of I, in order to be compatible with $\Diamond\alpha$, and, finally one for the equation itself. In this last case the self-referential ability is definitely disregarded. Therefore, we cannot consider the above equation (3) as a commonsense statement, although it has many advantages like, for example, providing an embedding for default reasoning in a wide class of modal logics (see [26–28,30,31,45,47]).

Let us reformulate the strict relation between definability and self-reference. As we have argued, classical definability is too strong because it is functional, i.e., it commits to uniqueness, and this behaviour is not reasonable in the real world. Weakening implicit definability (as nonmonotonic formalisms do) involves, in almost all cases, either a disjunction of definitions (in the case of circumscription) or a variety of extensions (e.g. in the case of default) which are, in general, infinite objects not characterizable through sentences. Each disjunct or extension implicitly refers to a different context. Insofar as the context is not made explicit in the language, an agent does not have, in general, an appropriate sentence for representing the required necessary and sufficient conditions involved in a definition.

What we need is a sentence whose denotation depends on its context, like the denotation of the word "here" depends on the place it is uttered [51]. A well-known device for naming a context is to treat it as a designator, i.e., a parameter which acquires its meaning through a suitable substitution (see [50] for a discussion on the role of substitution in these cases). For example, to express that a given sentence α is consistent relative to a context C a formula of the form $C \wedge \alpha$ may be provided, say $\varphi[C]$. Since C occurs in $\varphi[C]$ as a parameter, a suitable substitution for C has to be found. The designator C is, indeed, the very sentence $\varphi[C]$. Therefore the substitution for C is a sentence Δ, in which C does not occur, such that $\varphi[\Delta] \equiv \Delta$ is true. The sentence $\varphi[C] \equiv C$ is a fixed point equation which gives us explicit definitions for C: they are the admissible substitutions Δ for C in $\varphi[C]$, established by showing that $\varphi[\Delta] \equiv \Delta$ is a theorem.

More precisely, when a context is treated as a parameter in a self-referential language, finding a substitution for the context amounts to find an explicit definition for it. Such an explicit definition exists only if we can prove that there are theorems of the form $\vdash_{\Lambda} \varphi[\Delta] \equiv \Delta$, where Λ is an appropriate logic that will be made precise in the paper.

The above considerations imply that an explicit definition of a context requires a logic where fixed point theorems of the form $\vdash_{\Lambda} \varphi[\Delta] \equiv \Delta$ exist for a formula $\varphi[C]$ in which the parameter C is suitably used as a designator. More important, the existence of such fixed point theorems must be stated without resorting to the uniqueness of C, that is, without relying on an implicit definition of C. This has been, in fact, the main effort of this work. The connection between nonmonotonic reasoning and a logic in which fixed points are characterizable, namely the modal logic G, has been early investigated by [8]. The modal logic G [7,52] is the logic in which the notion of *provability* in Peano Arithmetic is interpreted. For this reason the characterization of fixed points goes through implicit definability which, as we discussed above, means that there exists a unique solution to the fixed point equation; this commitment to uniqueness brings us back to the constraints of first-order logic. The inadequacy of G for interpreting consistency and provability in nonmonotonic logic was noticed by Doyle, who also observed that both G and G^*—axiomatizing the notion of truth in Peano arithmetic—"miss out on all *contingent* statements of provability" [9]. We discuss this point in Section 4.2. On the other hand, Gabbay [14] has modeled negation by failure in logic programming by means of the provability operator of the modal logic G.

Let us consider again the previous example and say that $T[C]$ is a theory in which a context C occurs as a parameter. That is, $T[C]$ is a theory in which one can represent, e.g., sentences of the form:

$$(C \rightarrow (\forall x\, \mathrm{RED}(x) \rightarrow \mathrm{ONTABLE}(x)))$$

or

$$(C \wedge \mathrm{ONTABLE}(a)).$$

Then, what we want to get from $T(C)$ are the following explicit definitions:

$$\Delta_1 \equiv (\forall x\, \mathrm{ONTABLE}(x) \equiv \mathrm{RED}(x) \vee x = a),$$
$$\Delta_2 \equiv (\forall x\, \mathrm{ONTABLE}(x) \equiv \mathrm{RED}(x) \vee x = b).$$

Now, such sentences Δ_1 and Δ_2 exist if we can say that there is a logic Λ in which, by substituting either Δ_1 or Δ_2 for C in $T[C]$, we get the following theorems:

$$\vdash_{\Lambda} T(\Delta_1) \equiv \Delta_1,$$
$$\vdash_{\Lambda} T(\Delta_2) \equiv \Delta_2.$$

We ask, furthermore, that if such fixed points are expressible in the logic Λ then the Δ_i are computable from $T[C]$ itself.

The problem is to determine the restrictions that has to be imposed so as to avoid the obvious paradoxes of a self-referential language. Both in [37,38] and in [34] the difficulties of dealing with a first-order self-referential language are thoroughly analyzed. In particular Montague points out the problems concerned with substitution; we discuss this point in the last section of this paper.

It is possible to circumvent the difficulties arising from self-reference through a modal logic; this was in fact the claim of Montague. We show that there is a suitable modal

logic in which fixed points of predicates are definable, so that self-reference can be fully managed in the previously described sense. The modal approach to the problem is, therefore, just an initial way to inquire into interesting solutions. The main contribution of this paper is the introduction of a self-referential language, through modal logic, in which commonsense statements are expressible and in which explicit definitions are obtainable via fixed point equations.

More precisely, we show how one can deal with a self-referential language and that the context is, indeed, what we have in mind: in the case of a nonmonotonic theory which is implicitly defining a concept, the context is its explicit definition and, more generally, the context is the minimal set of formulae whose consequences are true in the theory. To show this, let us consider a default rule $\frac{\alpha:\beta}{\gamma}$ in default logic, let C be a parameter denoting the context and consider the following sentences:

(i) "α is provable with respect to the context C" is identified with $\Box(C \rightarrow \alpha)$.

(ii) "β is consistent with respect to the context C" is identified with $\Box \Diamond (C \wedge \beta)$.

Observe that we use the modal operator $\Box$ to interpret the notion of provability and the composite operator $\Box\Diamond$ to interpret the notion of consistency.

There are many logics in which consistency and provability are not dual, likewise $\Box$ is not always the dual of $\Diamond$—just consider intuitionistic logic [3]. Here we are manufacturing the meaning of the notions of consistency and provability in nonmonotonic logic. Since these notions are interpreted with respect to a context, they do not enjoy duality as in classical logic.

We define a sentence $E(C)$ capturing (i) and (ii) and such that there exists a diagonal sentence Δ, not containing C, that explicitly defines C in Λ, with Λ the modal logic $KD4Z$, i.e., the modal logic built from the axiom schemata K, D, 4 and Z,

$$\vdash_{KD4Z} \Delta \equiv E(\Delta). \tag{4}$$

The above fixed point equation provides a generalization, which correctly captures the relativized context, of explicit definability for theories expressed in this logical framework, where these theories may be used to define concepts. Specifically, when we apply this result to default theories we show that, whenever $\langle W, D\rangle$ is a default theory [39]:

$$\vdash_{KD4Z} \Delta \equiv E(\Delta) \tag{5}$$

iff

Δ^* provides an extension of the default theory $\langle W, D\rangle$

where * is an effective mapping from the language of Λ to that of $\langle W, D\rangle$.[3]

Example 1. Let us consider again the foregoing example, in which the default theory $\langle W, D\rangle$ is used to weakly implicitly define the predicate ONTABLE:

$$W = \{\forall x\, \mathrm{RED}(x) \rightarrow \mathrm{ONTABLE}(x), \mathrm{ONTABLE}(a) \vee \mathrm{ONTABLE}(b)\},$$

[3] By adding to $KD4Z$ the modal schema 5, the above result does no longer hold.

$$D = \left\{ \frac{: \neg \text{ONTABLE}(a)}{\text{ONTABLE}(b)}; \frac{: \neg \text{ONTABLE}(b)}{\text{ONTABLE}(a)}; \frac{: \text{ONTABLE}(x) \rightarrow \text{RED}(x)}{\text{ONTABLE}(x) \rightarrow \text{RED}(x)} \right\}.$$

Now, let $\langle W, D \rangle(C)$ be the self-referential sentence, obtained from $\langle W, D \rangle$, by making the context C explicit, as a designator. According to the results of this paper, we will have the following sentence, along the lines of (i) and (ii) introduced above:

$$\begin{aligned}
&[\Diamond C \wedge \Box(\forall x\, \text{RED}(x) \rightarrow \text{ONTABLE}(x) \wedge (\text{ONTABLE}(a) \vee \text{ONTABLE}(b))) \wedge \\
&\ [\Box(C \rightarrow \top) \wedge \Box \Diamond (C \wedge \neg \text{ONTABLE}(a)) \rightarrow \Box(C \rightarrow \text{ONTABLE}(b))] \wedge \\
&\ [\Box(C \rightarrow \top) \wedge \Box \Diamond (C \wedge \neg \text{ONTABLE}(b)) \rightarrow \Box(C \rightarrow \text{ONTABLE}(a))] \wedge \\
&\ [\Box(C \rightarrow \top) \wedge \Box \Diamond (C \wedge \forall x(\text{ONTABLE}(x) \rightarrow \text{RED}(x))) \rightarrow \\
&\qquad \Box(C \rightarrow (\forall x\, \text{ONTABLE}(x) \rightarrow \text{RED}(x)))]].
\end{aligned}$$

The technical meaning of this self-referential sentence will be made clear in the paper. Let us call $E(C)$ the boxed sentence $\Box\langle W, D \rangle(C)$. By suitably treating the parameter C it is possible to prove, using the techniques of this paper, that there are two sentences, namely

$$\begin{aligned}
\Delta_1 \equiv \Box\Box\, &(\text{RED}(b) \wedge (\text{ONTABLE}(a) \vee \text{ONTABLE}(b)) \wedge \\
&(\forall x\, \text{RED}(x) \rightarrow \text{ONTABLE}(x)) \wedge \\
&(\forall x\, \text{ONTABLE}(x) \rightarrow \text{RED}(x))) \wedge q,
\end{aligned}$$

and

$$\begin{aligned}
\Delta_2 \equiv \Box\Box\, &(\text{RED}(a) \wedge (\text{ONTABLE}(a) \vee \text{ONTABLE}(b)) \wedge \\
&(\forall x\, \text{RED}(x) \rightarrow \text{ONTABLE}(x)) \wedge \\
&(\forall x\, \text{ONTABLE}(x) \rightarrow \text{RED}(x))) \wedge q',
\end{aligned}$$

such that

$$\vdash_{KD4Z} E(\Delta_i) \equiv \Delta_i$$

for $i = 1, 2$ where q and q' in Δ_1 and Δ_2 are the "consistency part" of the context, i.e., are of the form $\Box \Diamond \alpha$, and are later eliminated by theoremhood, thus delivering Δ_i free from subformulae denoting consistency, and Δ_1 and Δ_2 are, respectively, substituted for C in $E(C)$. In other words, both Δ_1 and Δ_2 are equivalent to C, which is obtained from the fact that both Δ_1 and Δ_2 are fixed points of E in the modal logic $KD4Z$.

By mapping the formula Δ_1 into the language of $\langle W, D \rangle$ we get

$$\Delta_1^* = (\text{ONTABLE}(b) \wedge \forall x(\text{RED}(x) \equiv \text{ONTABLE}(x))) \tag{6}$$

which, in its turn, is equivalent to

$$\forall x(\text{ONTABLE}(x) \equiv (\text{RED}(x) \vee x = b)) \wedge \text{RED}(b). \tag{7}$$

Therefore we get an explicit definition out of the default theory. Observe that the sentence we have obtained through the fixed point is free from disjunctions.

The rest of the paper is organized as follows. In the next section we give some preliminaries on the language. In Section 3 we present the modal logic $KD4Z$ and a

tableau method to perform proofs. More details can be found in [1,2]. In Section 4 we present the results on definability of fixed points in $KD4Z$, we discuss self-reference and fixed points in modal logic, notably the relation with the provability logic G and extensions to quantified modal logic. In Section 5 we prove that Reiter extensions can be computed by a self-referential sentence in $KD4Z$ and some examples are provided in Section 6. Section 7 closes the presentation with a discussion and comparison with modal approaches to nonmonotonic reasoning. Finally we add an appendix with the proofs of the main theorems. Other proofs can be found in [1,2].

2. Preliminaries

We deal with propositional modal logic, we shall address the extension to first-order logic at the end of Section 4. We refer the reader to [17] for the basics of modal logic. A propositional modal language $\mathcal{L}$ is defined using a set of propositional letters Π whose elements are denoted $p, q, \ldots$ and a unary operator $\Box$. The well-formed formulae of the modal language are given by the rule

$$A ::= p \mid \bot \mid \neg A \mid A_1 \wedge A_2 \mid \Box A$$

where p ranges over elements of Π, $\bot$ denotes a contraddiction, $\top$ a tautology and the usual classical abbreviations for disjunction and implication apply. We use both lower-case Greek letters and upper-case Latin letters to denote sentences and reserve the last upper-case letters of the Latin alphabet and upper-case Greek letters to denote sets. Structures are denoted by Gothic letters.

In general, if $\Box$ is interpreted in the standard way as "necessity" then the modal context is *alethic* (from the Greek "true"). There are many other interpretations, like *epistemic*, where $\Box$ is interpreted either as "knowledge" or "belief", or *temporal*, where $\Box$ is interpreted as "always". Other meanings of $\Box$ are the *dynamic* one, that is, "true after every execution of an action", the *deontic* one, in which $\Box$ means "ought to be" and finally the *default* one, where a suitable interpretation of $\Box$ is "it is provable", similarly to the logic of arithmetic. This is, indeed, the intended meaning of the modal operator in this paper.

A dual operator of $\Box$ is defined by $\Diamond A = \neg \Box \neg A$.

We first introduce some basic notions on Kripke semantics for modal logic. A *Kripke frame* is a pair $\mathfrak{F} = \langle W, R \rangle$ where W is a nonempty set and R is a binary relation on W. A *model* is a pair $\mathfrak{A} = \langle \mathfrak{F}, I \rangle$, where $\mathfrak{F}$ is a frame, and I is a function assigning a subset $I(p)$ of W to each propositional letter p. The function I is called a *valuation*. The notation $\mathfrak{A}, w \models \varphi$ is defined inductively:

$$\mathfrak{A}, w \models p \quad \text{iff} \quad w \in I(p),$$

$$\mathfrak{A}, w \models \neg\varphi \quad \text{iff} \quad \text{not } \mathfrak{A}, w \models \varphi,$$

$$\mathfrak{A}, w \models \varphi \wedge \psi \quad \text{iff} \quad \mathfrak{A}, w \models \varphi \text{ and } \mathfrak{A}, w \models \psi,$$

$$\mathfrak{A}, w \models \Box\varphi \quad \text{iff} \quad \text{for all } v \in W, \text{ with } wRv, \mathfrak{A}, v \models \varphi.$$

If $\mathfrak{A} = \langle W, R, I \rangle$ is a model then $\mathfrak{A}, w \models A$ means that the formula A is *true* at world (point) w in the model $\mathfrak{A}$; A is *true* in a model $\mathfrak{A}$ ($\mathfrak{A} \models A$) if it is true at all worlds in $\mathfrak{A}$. A modal formula A is valid on a frame $\mathfrak{F}$ if A is true at every world of $\mathfrak{F}$, under every valuation.

Most of the research in commonsense reasoning dealing with knowledge, belief and self-reference takes into account *normal* modal logics (e.g. [28, 30, 36, 38, 42, 53, 55]). A *normal* modal logic is a collection of well-formed formulae of the modal language $\mathcal{L}$ that extends propositional logic with the axiom schema

$$K: \quad \Box A \wedge \Box(A \rightarrow B) \rightarrow \Box B$$

and it is closed under the rules of:

$$\text{Necessitation}: \quad A / \Box A;$$
$$\text{Uniform Substitution}: \quad A[p]/A[p/B],$$

where $p, A, B \in \mathcal{L}$, p is atomic and B is *uniformly* substituted for p in A. The resulting formula $A[p/B]$ is a *substitution instance* of A.

The following list includes some of the better known axiom schemata in modal logic, together with their traditional names:

D:	$\Box\varphi \rightarrow \Diamond\varphi$	(seriality),
4:	$\Box\varphi \rightarrow \Box\Box\varphi$	(transitivity),
T:	$\Box\varphi \rightarrow \varphi$	(reflexivity),
5:	$\Diamond\Box\varphi \rightarrow \Box\varphi$	(Euclideanness),
B:	$\Diamond\Box\varphi \rightarrow \varphi$	(symmetry),
Z:	$\Box(\Box\varphi \rightarrow \varphi) \rightarrow (\Diamond\Box\varphi \rightarrow \Box\varphi)$	(discreteness),
L:	$\Box(\Box\varphi \wedge \varphi \rightarrow \psi) \vee \Box(\Box\psi \wedge \psi \rightarrow \varphi)$	(linearity),
F:	$(\Diamond\Box\varphi \wedge \psi) \rightarrow \Box(\varphi \vee \Diamond\psi)$,	
X:	$\Box\Box\varphi \rightarrow \Box\varphi$	(density),
W:	$\Box(\Box\varphi \rightarrow \varphi) \rightarrow \Box\varphi$	(finiteness).

If $S_1, \ldots, S_n$ are schemata then $KS_1 \ldots S_n$ is the normal modal logic generated by $S_1, \ldots, S_n$. A sentence A is *provable* in the logic $\Lambda = KS_1 \ldots S_m$, denoted $\vdash_\Lambda A$ if it has a proof from $K, S_1, \ldots, S_m$. That is, there is a sequence of formulae $A_0, \ldots, A_n = A$, and for each i, $i \leqslant n$, either A_i is a propositional tautology or is an instance of $K, S_1, \ldots, S_m$, or it has been obtained by Modus Ponens or by Necessitation. The set of theorems of Λ coincides with the set of formulae in Λ that have a proof from $K, S_1, \ldots, S_m$.

Let $\mathbb{F}$ be a class of frames (or models), a normal modal logic Λ is *sound* with respect to $\mathbb{F}$ iff for all formulae φ and all $\mathfrak{A}$, $\mathfrak{A} \in \mathbb{F}$, $\vdash_\Lambda \varphi$ implies $\mathfrak{A} \models \varphi$. If Λ is sound with respect to $\mathbb{F}$ then $\mathbb{F}$ is said to be a *class of frames* (or models) for Λ.

A normal modal logic Λ is *complete* with respect to $\mathbb{F}$ iff for any set of formulae $\Gamma \cup \{\varphi\}$, if $\Gamma \models_{\mathbb{F}} \varphi$ then $\Gamma \vdash_\Lambda \varphi$. Hence Λ is said to be characterized by $\mathbb{F}$. E.g. KD is characterized by the class of *serial* frames, $S5$ is characterized by the class of *universal* frames, i.e., $R = W \times W$.

3. The modal logic *KD4Z*

The modal logic we are interested in is *KD4Z*, which is the normal modal logic extending *KD4* (i.e., the normal modal logic with transitivity and seriality) with the axiom *Z*, *discreteness*:

$$Z: \quad \Box(\Box\varphi \rightarrow \varphi) \rightarrow (\Diamond \Box \varphi \rightarrow \Box\varphi).$$

In the sequel we show that *KD4Z* is indeed a good logic for self-reference in commonsense reasoning. We shall illustrate the logic and give both semantics and proof theoretic methods. The schema *Z* is widely discussed in [17], more details can be found in [1,2].

3.1. *KD4Z and its semantics*

We now introduce some useful notions. Given a set T of worlds, with $u, v \in T$, if uRv and for no t is $uRtRv$ then u is called *predecessor* of v and v *successor* of u. Note that both a successor or a predecessor of a world may not be unique and a reflexive world has no predecessor or successor at all. A world $w \in T$ is called a *first element* in T if it has no preceding elements.

A frame is *linear* or *connected* if for all u and v either uRv or vRu; an irreflexive frame (i.e., for no u is uRu) is *weakly discrete* if for all worlds u and v, with uRv no infinite chains of worlds $uRt_1Rt_2 \ldots Rt_nR \ldots Rv$ exist; a frame is *universal* if for all u and v, uRv. We recall from [28] that the *concatenation* $R_1 \odot R_2$ of two relations $R_1 \subseteq T \times T$ and $R_2 \subseteq S \times S$ is the relation $R_1 \cup (T \times S) \cup R_2$.

Given two Kripke frames $\mathfrak{F}_1 = \langle T, R_1 \rangle$ and $\mathfrak{F}_2 = \langle S, R_2 \rangle$, where $T \cap S = \emptyset$, the Kripke frame $\langle T \cup S, R_1 \odot R_2 \rangle$ is called the *concatenation* of $\mathfrak{F}_1$ and $\mathfrak{F}_2$ and is denoted by $\mathfrak{F}_1 \odot \mathfrak{F}_2$ (see[28]).

In a frame $\langle W, R \rangle$ a *cluster* is a subset V of W that is maximal with respect to the property that for all u and v from V, uRv. We say that a frame has the *terminal cluster* property if it is the concatenation of $\mathfrak{F}_1 \odot \mathfrak{F}_2$ where $\mathfrak{F}_2$ is a universal frame (compare with condition 2 of Definition 9.17 for *cluster-closed* class of models in [28]).

A frame $\mathfrak{F}$ is said to be *well-capped* or to have the *finite depth property* if no ascending chain of worlds $w_1R \ldots Rw_nRw_{n+1} \ldots$ exists. A frame $\mathfrak{F}$ is said to be of *depth n* if no $(n+1)$-chain of worlds $w_1R \ldots Rw_nRw_{n+1}$ exists.

The following theorem gives the characterizations of *KD4Z*.

Theorem 2 (see [1]). *KD4Z is complete with respect to each of the following*:

- *The class* $\mathfrak{G}_{fin}$ *of finite models whose frames are* $\mathfrak{F} \odot V$, *where* $\mathfrak{F}$ *is transitive, irreflexive with a first element and V is a cluster.*
- *The class* $\mathfrak{G}$ *of models whose frames are* $\mathfrak{F} \odot V$, *where* $\mathfrak{F}$ *is transitive, irreflexive, well-capped and V is a cluster.*
- *The class* $\mathfrak{G}$ *of models whose frames are* $\mathfrak{F} \odot V$, *where* $\mathfrak{F}$ *is transitive, irreflexive, well-capped, V is isomorphic to* $\langle \omega, < \rangle$.

The decidability of *KD4Z* follows from the above theorem.

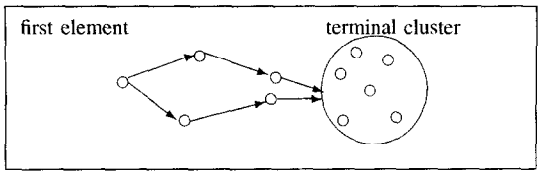

Fig. 1. A $KD4Z$ model.

By adding the linearity axiom L to $KD4Z$ we have that $KD4LZ$ is complete with respect to the class $\mathcal{B}$ of *balloons* [17], where the frames are the results of the concatenation of finite, linear, irreflexive and with a first element frames with a terminal cluster. $KD4LZ$ is also complete with respect to the frame of "integers" $\langle \omega, < \rangle$.

Note that a frame with no ascending chain is not necessarily finite: in fact a world w may have an infinite number of successors (or predecessors) w_i, $i < \delta$, where δ is any cardinal. The set of all successors of a world is called a *pseudo-cluster*.

Fig. 1 gives an intuitive picture of the structure of the frames for $KD4Z$.

3.2. Tableaux for KD4Z

We introduce in the following semantic tableaux for the logic $KD4Z$, a proof method useful when dealing with examples involving some derivation in the logic.

Semantic tableaux are used as refutation systems. They are built by means of a set of rules that preserve satisfiability of sets of formulae. We make use of the signed version of modal tableaux given in [13]. A formula φ is signed if it is presented in the form $T\varphi$ or $F\varphi$, where the prefixes T and F intuitively refer to *true* and *false*.

Each world in a $KD4Z$ model $\langle \mathfrak{F}, R, I \rangle$ can "see" a world belonging to the terminal cluster V and, as tableaux build a countermodel, we have to be able to end up in the cluster, whichever path we are following, after a given number of steps.

In the following the *possible force* formulae are signed formulae of the form $T \Diamond \alpha$, $F \Box \beta$, and the *necessity force* formulae are signed formulae of the form $T \Box \alpha$, $F \Diamond \beta$.

Modal expansion rules are stated as follows, where set union is briefly denoted by the comma, in the obvious sense.

3.3. Tableau rules

Let

$$\Gamma^{\sharp} = \{F \Box \Diamond p \mid F \Box \Diamond p \in \Gamma\} \cup \{T \Diamond \Box p \mid T \Diamond \Box p \in \Gamma\} \cup \{Tp, T \Box p \mid T \Box p \in \Gamma\} \cup \{Fp, F \Diamond p \mid F \Diamond p \in \Gamma\}.$$

Let

$$\Gamma^{\pi} = \{T \Diamond p \mid T \Diamond p \in \Gamma\} \cup \{F \Box p \mid F \Box p \in \Gamma\}.$$

$$D: \quad \frac{\Gamma}{\Gamma^{\natural}},$$

$$T\Box: \quad \frac{T\,\Box\varphi}{T\,\Diamond\,\Box\varphi},$$

$$F\Diamond: \quad \frac{F\,\Diamond\,\varphi}{F\,\Box\,\Diamond\varphi},$$

$$T\Diamond: \quad \frac{\Gamma, T\,\Diamond\,\varphi}{\Gamma^{\natural}, \Gamma^{\pi}, T\,\Box\,\Diamond\varphi, T\varphi \mid \Gamma^{\natural}, F\,\Diamond\,\varphi, T\varphi},$$

$$F\Box: \quad \frac{\Gamma, F\,\Box\,\varphi}{\Gamma^{\natural}, \Gamma^{\pi}, F\,\Diamond\,\Box\varphi, F\varphi \mid \Gamma^{\natural}, T\,\Box\,\varphi, F\varphi}.$$

The above rules for $F\Box$) and $T\Diamond$) are obtained from the axiom Z in the form $\neg\Box\alpha \rightarrow \Diamond(\Box\alpha \wedge \neg\alpha) \vee \neg\Diamond\Box\alpha$ together with the $KD4Z$ theorem $\Box\Diamond\alpha \rightarrow \Box\Box\alpha \wedge \Diamond\alpha$ capturing both transitivity and seriality.

If Γ is a set of signed formulae, a $KD4Z$ tableau for Γ is a tree whose root is labeled by Γ and every non-root node is obtained from a preceding node in the same branch by means of the application of an expansion rule of the logic. A branch $\mathcal{B}$ of a tableau is *closed* if both $T\varphi$ and $F\varphi$ occur in some node of $\mathcal{B}$, for some signed formula φ, or if $T\bot$ or $F\top$ occurs in $\mathcal{B}$; otherwise it is called *open*. A branch $\mathcal{B}$ is satisfiable if there exist a model and a world w such that w satisfies the conjunction of all the formulae occurring in $\mathcal{B}$. A tableau is satisfiable if some branch of the tableau is satisfiable. An open branch is satisfiable, a closed branch is unsatisfiable.

A tableau is closed iff all its branches are closed, i.e., if all its branches are unsatisfiable. A $KD4Z$ *tableau refutation* of Γ is a closed $KD4Z$ tableau for Γ. A tableau refutation of $F\varphi$ is a *tableau proof* of φ. Hence by a $KD4Z$ tableau proof for φ we mean a closed $KD4Z$ tableau whose root node is labeled by $F\varphi$.

Theorem 3 (Soundness). *If φ has a $KD4Z$ tableau proof then φ is valid in all $KD4Z$ models.*

Theorem 4 (Completeness). *If φ is valid in all the $KD4Z$ models, then φ has a $KD4Z$ tableau proof.*

The above theorems have fairly classical proofs along the lines of [13]; in particular, completeness exploits both filtration and the construction of a generated model. For details see [2].

4. Self-reference in modal logic

In the following we give the technical details on the definability of fixed points in the modal logic $KD4Z$; we shall discuss this result and compare it with those established in

provability logic G in the next section. We shall also address the problem of extensions to quantified modal logic.

4.1. Definability of fixed points in KD4Z

A modal sentence with occurrences of a propositional variable p will be denoted by $B(p)$ or $C(p)$, while $E(p)$ denotes any modal sentence where all occurrences of the propositional variable p are in the scope of some modal operator. $E(p)$ is called a *boxed* sentence. We may use $B(\Box C(p))$ instead of $E(p)$ to denote that $E(p)$ is the result of uniformly substituting $\Box C(p)$ for q in $B(q)$. By $E(A)$ we denote the formula obtained by uniformly substituting a formula A for p in $E(p)$. For example if $E(p)$ is $\Box(p \rightarrow \gamma)$ and p denotes the theory $\Box(\alpha \wedge \beta)$ then the substitution of $\Box(\alpha \wedge \beta)$ for p in $E(p) = \Box(p \rightarrow \gamma)$ gives $\Box(\Box(\alpha \wedge \beta) \rightarrow \gamma)$. p is, thus, used as a parametric theory, i.e., p denotes the theory which, substituted for p in $E(p)$, returns a formula in which the theory itself is realised.

A *fixed point* A for $E(p)$ in a logic Λ is a modal formula A logically equivalent to $E(A)$, that is $\vdash_{\Lambda} A \equiv E(A)$.

$\top$ denotes any tautology.

We say that

(i) a set of *fixed points* Φ explicitly defines a modal predicate E in $KD4Z$ provided

$$\vdash_{KD4Z} E(A) \equiv A \quad \text{iff} \quad A \in \Phi, \tag{8}$$

(ii) $E(p)$ is an *implicit consistency predicate* provided

$$\vdash_{KD4Z} E(p) \rightarrow \Box \Diamond p. \tag{9}$$

By suitably modifying some of the proofs reported in [48], we are able to prove that any implicit consistency predicate is definable in $KD4Z$. Furthermore, for a predicate of the form $\Box C(p)$ we are able to characterize, up to logical equivalences, all fixed points in $KD4Z$.

Theorem 5 (Definability of predicates of the form $\Box C(p)$). *Let $E(p) = \Box C(p)$, with $E(p)$ an implicit consistency predicate. Then $\Box C(\top)$ is a fixed point. That is,*

$$\vdash_{KD4Z} \Box C(\top) \equiv \Box C(\Box C(\top)).$$

The above theorem, whose proof is in Appendix A, states the existence of consistent solutions. That is, any implicit consistency predicate has a fixed point. The above condition (9), is essential to weaken implicit definability. In other words, the effect of this condition in the logic $KD4Z$ is to withdraw uniqueness of fixed points (see the discussion in the next section).

Example 6. Let $E(p) = \Box C(p) = \Box((\Box \Diamond (p \wedge A) \rightarrow \Box(p \rightarrow A)) \wedge \Diamond p)$. $E(p)$ is an implicit consistency predicate. $\Box C(\top)$ is $\Box(\Box \Diamond A \rightarrow \Box A)$.

We now provide a representation theorem for fixed points in the restricted hypothesis that $E(p)$ is of the form $\Box C(p)$. First, the fixed point $\Box C(\top)$ is the top element in the lattice of the fixed points of $E(p)$, while $\bot$ is the bottom element (being $\vdash_{KD4Z} \bot \equiv \Box \Diamond \bot \wedge \varphi$). In fact:

Theorem 7 (Existence of the maximum). *Every fixed point $\mathcal{T}$ of $\Box C(p)$ is of the form $\Box C(\top) \wedge \gamma$ for some γ, that is $\vdash_{KD4Z} \mathcal{T} \rightarrow \Box C(\top)$.*

In conclusion, we have:

Theorem 8 (Characterization of the structure of fixed points). *$\mathcal{T}$ is a fixed point of $E(p) = \Box C(p)$, with $E(p)$ an implicit consistent predicate, if and only if $\mathcal{T}$ is logically equivalent to a formula of the form $\Box C(\top) \wedge \Box \Diamond \gamma$.*

The above theorems, whose proofs can be found in Appendix A, show that whenever the language is finite then a finite set Φ of fixed points, of an implicit consistency predicate $E(p)$, is generated. Of course Φ is finite, up to logical equivalence. Furthermore Φ is partially ordered by the $\vdash_{KD4Z}$ relation as follows:

$$B \leqslant A \quad \text{iff} \quad \vdash_{KD4Z} B \rightarrow A. \tag{10}$$

Example 9. Let $E(p) = \Box((\Box \Diamond (p \wedge A) \rightarrow \Box(p \rightarrow A)) \wedge \Diamond p)$, as in Example 6. Then, $\Diamond \Box \neg A, \Box \Box A, \Box(\Box \Diamond A \rightarrow \Box A), \bot \in \Phi$. Observe also that $\Box(\Box \Diamond A \rightarrow \Box A)$ is the top element and $\bot$ is the bottom element, with respect to the ordering induced by (10).

It is also worth noting that elements in Φ may be orthogonal that is: $\vdash_{KD4Z} \neg(A_1 \wedge A_2)$ and $A_1, A_2 \in \Phi$.

4.2. A discussion on provability, consistency and the logic G

There is a well-established area of pure logic where modalities have been used to interpret the notions of "provable" and "consistent" in *PA*, the first-order Peano arithmetic [4, 7, 48, 52]. Following Gödel's procedure for numbering theorems of *PA*, a first-order "provability" predicate can be constructed. Once the unary predicate of provability is interpreted as a modal operator $\Box$, according to Solovay's translation, the modal counterpart of the Diagonalization lemma holds for a restricted set of modal formulae as stated by the following condition:

p obeys the diagonalization restriction (DR) in $E(p)$ iff p is boxed in $E(p)$.

Smorynski, in [48], postulates the conditions for the existence of fixed points by introducing suitable extensions of the basic modal logic $K4$:

- DOL is the extension of $K4$ such that for each formula $E(p, q_1, \ldots, q_n)$, p obeys DR and a new operator $\delta_E(q_1, \ldots, q_n)$ is added together with the following axiom schema:

$$\delta_E(B_1,\ldots,B_n) \equiv E(\delta_E(B_1,\ldots,B_n),B_1,\ldots,B_n). \tag{11}$$

- DIL is the extension of $K4$ obtained by adding the following diagonalization rule:

$$\text{DiR:}\quad \frac{[s]((E(p) \equiv p) \rightarrow A)}{A}$$

where p obeys DR, it does not occur in A, and $[s]\alpha = \Box\alpha \wedge \alpha$.

- G is the extension of $K4$ with the axiom:

$$\text{Löb:}\quad \Box(\Box\alpha \rightarrow \alpha) \rightarrow \Box\alpha.$$

On the basis of the above extensions the following holds:

$$\vdash_{DOL} A \quad \text{iff} \quad \vdash_{DIL} A \quad \text{iff} \quad \vdash_G A.$$

The proof of the above result (see [48]) amounts to the proof that fixed points are implicitly and explicitly definable in G. In fact, denoting $H(p) = [s](E(p) \equiv p)$, the following holds.

- *Implicit definability (ID) of fixed points in G*:

$$\text{ID:}\quad \vdash_G H(p) \wedge H(q) \rightarrow (p \equiv q).$$

In addition, from ID, by the Beth theorem and DiR (both holding in G) it follows:

- *Explicit definability (ED) of fixed points in G*: there exists a sentence Δ containing all the variables of $E(p)$ other than p and such that:

$$\begin{aligned} &\vdash_G H(p) \rightarrow (p \equiv \Delta),\\ &\vdash_G \Delta \equiv E(\Delta). \end{aligned} \tag{12}$$

The role of implicit definability (ID) is to state uniqueness of fixed points which is crucial for the Solovay's first completeness theorem [52], showing that G is the logic of provability, i.e., a modal sentence is a theorem of G if all its translations are theorems of PA. On the other hand, for this very reason of uniqueness, it is not possible to state, from a theory in the logic G, that "a sentence p is consistent", i.e., $\Diamond p$, because it would lead to the inconsistency of G. Therefore, no possible self-reference to the context, required to mention consistency, can be carried out in G. Observe that, instead, this is always possible in $KD4Z$. The role played by the implicit consistency predicate is thus clear: in fact, for example, the Gödel sentence $\vdash_A \varphi \equiv \neg Pr(\lceil\varphi\rceil)$, in which $E(p) = \neg\Box p$ is not definable when A is $KD4Z$.

On the other hand the second completeness theorem of Solovay shows that G^* is the modal logic whose theorems are precisely those modal sentences of which all translations are true in the PA standard model, where G^* is obtained from G by dropping the necessitation rule and adding reflexivity, i.e., $\Box A \rightarrow A$. Therefore $\Diamond\top$ is a theorem of G^* while $\Box\Diamond\top$ is not [7,52].

The upshot is that in any logic in which PA is expressible the two schemata $\frac{\vdash A}{\vdash \Box A}$ and $\Box A \rightarrow A$ are incompatible, which has been in fact investigated by Montague [34]. These results should be compared with the use, in nonmonotonic modal logic, of the schema T together with the rule of necessitation. It seems clear that none of these logics obtained

by these additions can capture self-reference (see Section 7 for a discussion). On the other hand, the failure of uniqueness of fixed points in $KD4Z$ should be interpreted as a sign of a more expressive power than G. In fact, consistency can be expressed in $KD4Z$.

4.3. Self-reference in quantified modal logic

A natural question to be answered is whether, by extending the language to a first-order one, self-reference is still characterizable in the appropriate way. An answer to this question for the quantified modal logic G (QG) has been given by Montagna in [33] and by Smorynski in [49]. We shall discuss Smorynski's results on definability for QG and some of his counterexamples, and give a simple case for quantified $KD4Z$.

Let us preserve from the propositional calculus a propositional letter, say p, that we shall use as a parameter to name contexts, the propositional connectives, the truth values ($\top$ and $\bot$) and the modal operators. The quantified modal language $Q\mathcal{L}$ we are concerned with is obtained by adding to the above propositional constructs and symbols an infinite set of variables and n-ary predicate symbols, and the quantifiers $\forall$ and $\exists$. Let us call $QKD4Z$ (respectively QG) the extension of propositional $KD4Z$ (respectively G) with the instances of axioms of the predicate calculus in the above defined language.

Let us now add to $QKD4Z$ (respectively QG) the Barcan formula:

$$B: \quad \forall x \, \Box P(x) \rightarrow \Box \forall x \, P(x)$$

that is the syntactic counterpart of models with constant domains. The converse of the Barcan formula:

$$BC: \quad \Box \forall x \, P(x) \rightarrow \forall x \, \Box P(x)$$

is derivable in QG (see [49]) as well as in $QKD4Z$.

The difficulty of definability of fixed points in quantified modal logic stems mainly from the interplay between variables and modal operators. In the case quantifiers cannot be pushed against the $\Box$, notwithstanding the Barcan formula, there are counterexamples to the definability of fixed points. In Smorynski the following counterexample is given, among others:

$$A(p) = \forall x(\Box \Box P(x) \rightarrow \Box(p \rightarrow P(x))). \tag{13}$$

Eq. (13) is used by Smorynski, in particular, to show incompleteness for QG with respect to arithmetic interpretations. An example can be analogously used in $QKD4Z$ to show that if variables are not bound, with the Barcan formula and even with finite domains, definability fails.

We illustrate the claim as follows:

Example 10. Let $A(p) = \Diamond p \wedge \forall x(\Box \Box P(x) \rightarrow (p \rightarrow \Box P(x)))$. Define the terminal cluster as $\models P(0), P(1), P(2)$, $w_2 \models \neg P(0), P(1), P(2)$, $w_1 \models P(0), \neg P(1)$, $w_1 \models P(2)$, and $w_0 \models P(0), P(1), \neg P(2)$.

Then

$$w_2 \models \Box P(0), \Box P(1), \Box P(2), \Box\Box P(0), \Box\Box P(1), \Box\Box P(2),$$
$$w_1 \models \neg\Box P(0), \Box P(1), \Box P(2), \Box\Box P(0), \Box\Box P(1), \Box\Box P(2),$$
$$w_0 \models \neg\Box P(0), \neg\Box P(1), \Box P(2), \neg\Box\Box P(0), \Box\Box P(1), \Box\Box P(2).$$

Therefore $w_0 \not\models p$ where $p = \forall x(\Box\Box P(x) \rightarrow \Box P(x))$ and $w_0 \models p \equiv \bot$, while in w_0 the following are all satisfied:

$$\Diamond p \wedge \forall x(\Box\Box P(x) \rightarrow (p \rightarrow \Box P(x))),$$
$$\Diamond p \wedge \forall x(\Box\Box P(x) \rightarrow (\bot \rightarrow \Box P(x))),$$
$$\Diamond p \wedge \forall x(\Box\Box P(x) \rightarrow \top).$$

Smorynski discusses also the cases which, with the Barcan formula, do not offer counterexamples to definability. In particular, interpreting the multimodal logic SR_n (see [48, Chapter 4]) into $QG + B$ it is possible to show that for any propositional combination of formulae of the form $Q_1x_1 \ldots Q_kx_k\Box B$, in which quantifiers are pushed against the box, explicit definability can be given.

Analogously, we show that for some restricted class of formulae of the quantified modal language, the first substitution lemma holds (FSL, see Appendix A). Then, since the second substitution lemma (SSL, see Appendix A) can be obtained by means of pure syntactical manipulations from FSL, for this class of formulae definability can be given.

Let $E(p)$ be a formula of $QKD4Z$, decomposable as $E(\Box C_1(p), \ldots, \Box C_k(p))$, with p not occurring in $E(q_1, \ldots, q_k)$, E propositional in $q_1, \ldots, q_k$ and each $\Box C_i(p)$ either propositional or of the form $\Box(p \circ Q_1x_1, \ldots, Q_kx_k\varphi(x_1 \ldots x_k))$, with all the occurrences of variables in $\varphi(x_1 \ldots x_k)$ bound, no modal operators occurring in $\varphi(x_1 \ldots x_k)$, and $\circ \in \{\rightarrow, \wedge, \vee\}$.

Then:

(1) $\vdash_{QKD4Z} \Box(A \equiv B) \rightarrow (\Box(A \circ Q_1x_1, \ldots, Q_kx_k\varphi(x_1 \ldots x_k)))$
$\equiv (\Box(B \circ Q_1x_1, \ldots, Q_kx_k\varphi(x_1 \ldots x_k)))$ (tautology)

(2) $\vdash_{QKD4Z} \Box(A \equiv B) \rightarrow \bigwedge_{i \leqslant k} \Box C_i(A) \equiv \Box C_i(B)$ ((1), tautology)

(3) $\vdash_{QKD4Z} \Box(A \equiv B) \rightarrow E(\Box C_1(A), \ldots, \Box C_k(A))$
$\equiv E(\Box C_1(B), \ldots, \Box C_k(B))$ ((2), propositional substitution)

(4) $\vdash_{QKD4Z} \Box(A \equiv B) \rightarrow (E(A) \equiv E(B))$ ((3))

(5) $\vdash_{QKD4Z} \Box(A \equiv B) \wedge (A \equiv B) \rightarrow (E(A) \equiv E(B))$ ((4), tautology)

(6) $\vdash_{QKD4Z} [s](A \equiv B) \rightarrow (E(A) \equiv E(B))$ ((5), tautology)

Observe that if $E(p)$ is decomposable as above, obviously the Barcan formula is useless. Both the Barcan formula and its converse may be needed just to get the decomposition. Therefore Theorem 5, i.e., the definability of predicates of the form $\Box C(p)$, can be analogously given for formulae of $QKD4Z$ in the restricted form defined above.

Example 11. Consider the following cases:

- Let $E(p) = \Box(p \rightarrow \forall x(P(x) \rightarrow Q(x))) \wedge \Box(p \wedge P(a) \vee P(b))$; then $E(p) = E(\Box C_1(p), \Box C_2(p))$, where $E(q_1, q_2) = q_1 \wedge q_2$, $\Box C_1(p) = \Box(p \rightarrow (\forall x(P(x) \rightarrow Q(x))))$ and $\Box C_2(p) = \Box(p \wedge P(a) \vee P(b))$. $\Box C_1(\top) = \Box(\forall x(P(x) \rightarrow Q(x)))$, $\Box C_2(\top) = \Box(P(a) \vee P(b))$ and $E(\top) = \Box(\forall x(P(x) \rightarrow Q(x))) \wedge \Box(P(a) \vee P(b))$.
- Let $E(p) = \forall x \Box(p \wedge P(x))$, then, by B and its converse $E(p) \equiv \Box \forall x(p \wedge P(x))$, $E(p) = E(\Box C(p))$, where $\Box C(p) = \Box(p \wedge \forall x P(x))$.
- If $E(p) = \forall x(Q(x) \vee \Box(p \rightarrow P(x)))$, then it has no good decomposition.

We can observe that the example of the introduction is, in fact, analogous to the first case above.

5. Self-reference in nonmonotonic logic: the case of default logic

We have shown that fixed points are explicitly definable in $KD4Z$. In this section we show that a default theory yields the definition of a modal predicate that has fixed points corresponding to an effective translation of its extensions.

Let $\langle W, D \rangle$ be a default theory, where W is finitely axiomatizable and D is a finite set of default rules each of the form $\delta = \frac{\alpha:\beta}{\gamma}$; α is called the *prerequisite*, β the *justification* and γ the *conclusion* of the default. We assume the reader familiar with Reiter's default logic, we refer to [39] and to [28] for the basic formalisms and the most significant results. Let p be a propositional parameter, denoting the context. We define the translation of a default theory into the modal logic $KD4Z$ as follows:

$$tr_\delta(p) = \Box(p \rightarrow \alpha) \wedge \Box \Diamond (p \wedge \beta) \rightarrow \Box(p \rightarrow \gamma), \tag{14}$$

$$Tr_{\langle W,D \rangle}(p) = \Box \left(\Diamond p \wedge W \wedge \bigwedge_{\delta \in D} tr_\delta(p) \right). \tag{15}$$

Observe that $\Diamond p$, under the outermost $\Box$, commits the contest to be consistent.

In addition, we define a *stability* condition $St_{\langle W,D \rangle}$ which singles out the set J_D of justifications of the defaults in D maximally consistent with the parameter theory p as:

$$St_{\langle W,D \rangle}(p) = p \rightarrow \Box \bigwedge_{\beta \in J_D} (p \wedge \Diamond \beta \leftrightarrow p \wedge \Box \Diamond \beta). \tag{16}$$

In the following we shall drop the subscript both in $Tr_{\langle W,D \rangle}(p)$ and $St_{\langle W,D \rangle}(p)$ when no confusion arises.

The fixed point equation becomes therefore

$$p \equiv (Tr(p) \wedge St(p)) \tag{17}$$

which is equivalent to

$$(p \equiv Tr(p)) \wedge St(p) \tag{18}$$

by definition of $St(p)$.

In the sequel we sometimes identify $(Tr(p) \wedge St(p))$ with $E(p)$. We call the solutions to the fixed point equation (18) *saturated fixed points*. Namely, a saturated fixed point $\mathcal{T}$ satisfies:

$$\vdash_{KD4Z} (\mathcal{T} \equiv Tr(\mathcal{T})) \wedge St(\mathcal{T}). \tag{19}$$

Since p is not boxed in $St(p)$, the theorems of Section 4 cannot be applied directly to $E(p)$. Nevertheless, the fact that the fixed point equation is equivalent to $(p \equiv Tr(p)) \wedge St(p)$ allows the computation of fixed points, notwithstanding the fact that in the general case a formula may have different decompositions and the computation of $\mathcal{T}$ may be rather complex. The set of saturated fixed points is thus a subset of the set of fixed points of Tr. Furthermore, we may use the "guess and check" method for computing fixed points (similar to [28]): we first look for candidates $\mathcal{T}$ by exploiting Theorem 8 applied to $Tr(p)$ and then check whether $St(\mathcal{T})$ holds. This will be explained in more details below.

Once a theory $\mathcal{T}$ is uniformly substituted for p in the self-referential sentence $p \equiv E(p)$ the first sentence Tr checks whether there exists a succession of defaults whose modal translations are closed in $\mathcal{T}$; the second sentence St checks for the saturation of the application of the whole set of defaults D.

The fixed point equation (19) states that the logical content of $\mathcal{T}$ is exactly circumscribed by the two conditions Tr and St. We can now introduce both a provability and a consistency operator for default logic. We define the provability operator $Pr_{\mathcal{T}}(\varphi)$ for default logic as $\Box(\mathcal{T} \rightarrow \Box\varphi)$ and the consistency operator $Con_{\mathcal{T}}(\varphi)$ as $\Box(\mathcal{T} \wedge \Diamond\varphi)$. Note that a fixed point solution $\mathcal{T}$ for $E(p)$ is of "necessary force", that is $\vdash_{KD4Z} \mathcal{T} \rightarrow \Box\mathcal{T}$, then $\mathcal{T} \vdash_{KD4Z} Pr_{\mathcal{T}}(\varphi)$ if and only if $\mathcal{T} \vdash_{KD4Z} \Box\Box\varphi$.

In the following $\mathcal{T}^*$ denotes the set of modal-free formulae α such that $\mathcal{T} \vdash_{KD4Z} \Box\Box\alpha$.

Let J_D be the set of justifications β of the defaults of D. To show that the condition St imposes the saturation of the application of the defaults of D, we must prove that, for the maximal subset D' of D whose elements in $J_{D'}$ are consistent with $\mathcal{T}^*$, the consistency of the elements in $J_{D'}$ can be derived from $\mathcal{T}$. This leads to prove:

Theorem 12. *Let*

$$\vdash_{KD4Z} (\mathcal{T} \equiv Tr_{\langle W,D\rangle}(\mathcal{T})) \wedge St_{\langle W,D\rangle}(\mathcal{T})$$

and

$$D' = \left\{ \left\langle \frac{\alpha : \beta}{\gamma} \right\rangle \mid \beta \text{ consistent with } \mathcal{T}^* \right\}.$$

Then:

(a) (*Reduct of a fixed point*) $\vdash_{KD4Z} \mathcal{T} \equiv Tr_{\langle W,D'\rangle}(\mathcal{T}) \wedge \bigwedge_{\beta \in J_{D'}} \Box\Diamond(\mathcal{T} \wedge \beta)$.

(b) (*Completeness*) $\mathcal{T}^*$ *is a Reiter extension of* $\langle W, D\rangle$.

Theorem 12(a) whose proof is that of Proposition A.13 in Appendix A, provides a method for reducing the search of solutions for fixed point equations to the case where

$E(p)$ has the form $\Box C(p)$. It shows that D' is the *reduct of* D with respect to the context $\mathcal{T}^*$ [16,28]. Theorem 12(b), whose proof is that of Proposition A.14 in Appendix A, shows that D' consists of those defaults used for constructing the extension $\mathcal{T}^*$.

As for the other direction of Theorem 12(b), if E is a Reiter extension, then the modal formula $\Box E \wedge \Box \Diamond \{\beta \mid \beta \in J_D$ and β consistent with $E\}$ is a saturated fixed point.

Theorem 12, together with the results presented in the previous sections, tells us how to find saturated fixed points. The method consists of the following steps:

- Substitute $\top$ for p in $Tr_{\langle W,D\rangle}(p)$, thus getting $Tr_{\langle W,D\rangle}(\top)$.
- $\bot$ is a fixed point solution of $Tr_{\langle W,D\rangle}(p)$. If it is the only solution, then there are no consistent extensions.
- All fixed points of $Tr_{\langle W,D\rangle}(p)$ are obtained as logical conjunctions of $Tr_{\langle W,D\rangle}(\top)$ with $\Box \Diamond \beta$ (possibly obtaining $\bot$).
- Theorem 12 applies to obtain Reiter extensions. Since we are interested in the reduction of the fixed points to the modal-free language by means of the operator $*$, it is sufficient to consider only those β which are justifications of defaults of $\langle W, D\rangle$. A consistent saturated fixed point, if any, is obtained from a fixed point $\mathcal{T}$ of $Tr_{\langle W,D\rangle}(p)$ by putting in logical conjunction $Tr_{\langle W,D\rangle}(\top)$ with a certain set of formulae $\Box \Diamond \beta$ (β must be consistent with $\mathcal{T}^*$). If it is consistent and still equivalent to $\mathcal{T}$, then it corresponds to a consistent Reiter extension via the reduction performed by the operator $*$.

6. Examples

We now show some examples to illustrate how to check for fixed points for Reiter extensions, along the lines of the given results. Observe that we shall give all the paradigmatic examples, so that any other (e.g. the one given in the introduction) can be easily reduced to the following ones.

Example 13. Let us consider the default theory $\langle W, D\rangle = \{\emptyset; \langle \frac{:\alpha}{\neg\gamma}, \frac{:\gamma}{\neg\alpha}\rangle\}$, which has two extensions. We have:

$$p \equiv \Box \Diamond p \wedge \Box[(\Box \Diamond (p \wedge \alpha) \rightarrow \Box(p \rightarrow \neg\gamma)) \wedge (\Box \Diamond (p \wedge \gamma) \rightarrow \Box(p \rightarrow \neg\alpha))] \wedge St_{\langle W,D\rangle}(p).$$

$Tr_{\langle W,D\rangle}(\top)$ is $\Box(\Diamond\alpha \rightarrow \Box\neg\gamma) \wedge \Box(\Diamond\gamma \rightarrow \Box\neg\alpha)$. By assuming the first default constituting D' as in Theorem 12, we get a candidate for a saturated fixed point $\mathcal{T}_1$, $\Box \Diamond \alpha \wedge Tr_{\langle W,D\rangle}(\top)$ which is $\Box(\Box\neg\gamma \wedge \Diamond\alpha)$. This is a saturated fixed point since D' satisfies the condition of Theorem 12. Analogously, a second saturated fixed point $\mathcal{T}_2$ is $\Box(\Box\neg\alpha \wedge \Diamond\gamma)$. The two Reiter extensions E are $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$, namely the deductive closures of $\neg\gamma$ and $\neg\alpha$ respectively.

Example 14. Let us consider the default theory $\langle W, D\rangle = \{\emptyset; \langle \frac{\alpha:\top}{\gamma}\rangle\}$, which has only the trivial extension $\{\top\}$. We have:

$$p \equiv \Box \Diamond p \wedge \Box[(\Box(p \rightarrow \alpha) \wedge \Box \Diamond (p \wedge \top) \rightarrow \Box(p \rightarrow \gamma))] \wedge St_{\langle W,D\rangle}(p).$$

$Tr_{\langle W,D\rangle}(\top)$ is $\Box(\Box\alpha \rightarrow \Box\gamma)$. This is the only saturated fixed point up to the reduction operator $*$. Then the only Reiter extension $Tr_{\langle W,D\rangle}(\top)^*$ is equivalent to $\top$.

Example 15. The default theory $\langle W = \emptyset, D = \{\frac{:\beta}{\neg\beta}\}\rangle$ has no Reiter extension.

$E(\top) =_{KD4Z} \Box(\Box \Diamond \beta \wedge \Box\top \rightarrow \Box\neg\beta) =_{KD4Z} \Box(\Box \Diamond \beta \rightarrow \Box\neg\beta) =_{KD4Z} \Diamond \Box \neg\beta$ is the generating fixed point for E but

$$\nvdash_{KD4Z} St_{\langle W,D\rangle}(\Diamond \Box \neg\beta).$$

While $\langle W = \bot, D = \{\frac{:\beta}{\beta}\}\rangle$ has $\bot$ as Reiter extension.

And, in fact, $E(\top) =_{KD4Z} \bot$ is a fixed point of E and also satisfies the stability condition:

$$\vdash_{KD4Z} St_{\langle W,D\rangle}(\bot).$$

Example 16. We now show with a simple example, a computation of a fixed point.

Let be $W = \emptyset$ and $D = \{\frac{:\beta}{\beta}\}\rangle$ in $\langle W, D\rangle$. Then $E(p) = \Box(\Diamond p \wedge (\Box \Diamond (p \wedge \beta) \rightarrow \Box(p \rightarrow \beta)))$.

$$Tr_{\langle W,D\rangle}(\top) = (\Box \Diamond \beta \rightarrow \Box\beta).$$

$\Delta = \Box\Box\beta$ is such that

$$E(\Delta) = E(\Box\Box\beta) = \Box(\Diamond \Box\Box\beta \wedge (\Box \Diamond (\Box\Box\beta \wedge \beta) \rightarrow \Box(\Box\Box\beta \rightarrow \beta)))$$

since in $KD4Z$, $\vdash \Diamond \Box\alpha \rightarrow \Box \Diamond (\Box\Box\alpha \wedge \alpha)$ and $\vdash \Diamond \Box\Box\alpha \rightarrow \Diamond \Box\alpha$ the following is obtained also by (4): $\Box(\Diamond \Box\beta \wedge (\Diamond \Box\beta \rightarrow \Box(\Box\beta \rightarrow \beta)))$ hence, by the schema Z: $\Box(\Box\alpha \rightarrow \alpha) \rightarrow (\Diamond \Box\alpha \rightarrow \Box\alpha)$, we get

$$\Box(\Diamond \Box\beta \wedge (\Diamond \Box\beta \rightarrow \Box\beta)) =_{KD4Z} \Box\Box\beta =_{KD4Z} \Delta.$$

Moreover, $\vdash_{KD4Z} St_{\langle W,D\rangle}(\Box\Box\beta)$.

While, if $\Delta = \Diamond \Box \neg\beta$ then $\nvdash_{KD4Z} St_{\langle W,D\rangle}(\Diamond \Box \neg\beta)$ $(=_{KD4Z} \Diamond \Box \neg\beta \rightarrow \Box(\Diamond \Box \neg\beta \wedge \Diamond\beta \leftrightarrow \bot) =_{KD4Z} \Diamond \Box \neg\beta \rightarrow \Box\Box \neg\beta)$.

The above examples seem to bring lots of machinery into the computation of Reiter extensions. Observe however that these computations can be completely automated, since $KD4Z$ is decidable and tableaux for the logic are available.

7. Discussion on the related literature

We have argued that much of commonsense reasoning concerns the definition of concepts, and that this is, in fact, a central theme in nonmonotonic reasoning.

Nonmonotonic reasoning provides the sufficient conditions in the definition of concepts by relativizing it to the context, which accounts for some self-referential statement

such as "It is consistent to assume, in the context p, that α is $\beta_1, \ldots, \beta_k$". However, strong (functional) definability like in G or first-order logic cannot account for the above notions.

This form of relativization to the context is often carried out, in the literature on nonmonotonic reasoning, by binding commonsense statements to the belief set of an agent. So that the self-referential ability is lifted from the commonsense language to one in which two distinct (or even more) levels of reasoning are formulated.

These approaches are based on two paradigms, which we call the *preference paradigm* and, according to Marek and Truszczynski, the *negation as failure to prove* paradigm.

The preference paradigm [5, 18, 25], defines a preference relation among modal structures which are sets of classical interpretations. Preference criteria for default reasoning were introduced by Doyle in [10] and Etherington [12]. Lin and Shoham [25] use a bimodal logic to provide a semantical characterization of default logics by means of preferred modal models.

The fixed point paradigm is based on the following idea: "Find a solution (i.e., a *Λ-expansion*) to the equation $\Gamma = Cn_\Lambda(I \cup \{\Diamond\gamma \mid \neg\gamma \notin \Gamma\}$", where I is a translation of the defaults in the modal logic Λ.

In particular the above schema has been introduced by the early work of McDermott and Doyle (see [30, 31]), further developed by Stalnaker in an unpublished manuscript of 1980, appeared later in [53], with the introduction of the notion of *stable set*, and afterwards expanded by Moore in [35]. The connection with default logic has been proposed by the work of Konolige [19] and further developed in [27].

Marek and Truszczynski [26, 28, 55] have shown that with the negation as failure to prove paradigm a family of modal logics can be devised to capture default reasoning. In fact, Λ-expansions with Λ in this family of modal logics are stable sets of ground theories Γ which turn out to be Reiter extensions. Indeed, Marek, Truszczynski and Schwartz have established more general results showing that Doyle and McDermott's fixed point is so powerful to yield infinitely many nonequivalent nonmonotonic modal logics, by varying the choice of the underlying monotonic modal logic [45]. Among these $KD45$, $Sw5$, $S4f$, $S4.2$ and $S4.3$ have been widely studied and proposed as good candidates for representing knowledge and belief. A strong argument in favor of them is that they are maximal [43–46] and even the largest in their range [45] (i.e., in the class of modal logics which generate the same nonmonotonic modal logic).

Recently Amati et al. [1] introduced boxed nonmonotonic modal logics, that is logics enjoying an alternative fixed point construction; for these logics the schema (3) is redefined by means of boxed contexts. *Boxed expansions* are thus generated by a set equation which is called *boxed fixed point*. The underlying idea is to interpret the membership relation of a formula to the context $\mathcal{T}$ as a nonmonotonic provability operator.

In general, modal logic is used to this purpose, belief and knowledge are treated as two modalities, and the agent can also define autoepistemic truths, i.e., sentences containing occurrences of K and B that refer to the agent's knowledge itself. In what way K and B should be clearly distinguished is however a controversial matter.

For example, Schwartz and Truszczynski [47] consider the logics of minimal knowledge of Halpern and Moses [18] more suitable for describing knowledge sets of an

agent than the autoepistemic logic of Moore or Levesque's "only knowing" logics [20]. For example, if $\Box$ is a modality standing for either B or K and the agent's initial assumption is $I = \{\Box p \rightarrow p\}$, then p belongs to the belief set when using Moore's autoepistemic logic [35], while Halpern and Moses' logics reject p as a plausible conclusion. Analogously, if the agent's belief is just $\{\Box p\}$, then it is argued that p should be a plausible conclusion, but when $\Box$ is used as a belief operator, p may be actually false, even though believed. Indeed, an implicit requirement for a knowledge operator K is to avoid forming a knowledge set containing p whenever the agent's belief is $\{Kp \rightarrow p\}$, since this last axiom schema must in some way be considered tautological. This remark yields an apparent paradox if one compares it with the valid reasoning schemata of G, in which the modal operator can be regarded as an extremely strong form of the knowledge operator, to the extent that what is known by the system is effectively provable. Contrary to our intuition, from the validity of $\{\Box p \rightarrow p\}$ it follows that p is a theorem (by the Löb rule). As a consequence, theories containing sentences of the form $\neg \Box \varphi$ cannot be considered in G, thus denying the possibility of using G as a logical basis for any form of negative introspection. This fact is a consequence of the well-known result about the incompleteness of first-order logic in the language of arithmetic; provability and truth run on two different tracks. This limiting result causes unrecoverable drawbacks to the possibility of distinguishing beliefs from knowledge: if the agent's knowledge is intended as a set of true beliefs and a suitable notion of truth can be given only at the intentional or "external" level, then the question of which modal logics are appropriate to represent knowledge and beliefs may become an eternally arguable question.

Another approach, to maintaining a self-referential ability in the language, is taken by quoting sentences referring to themselves in the language. This amounts to dealing with some analogue of the Löb derivability conditions like $Bel(\lceil\varphi\rceil) \wedge Bel(\lceil\varphi \rightarrow \psi\rceil) \rightarrow Bel(\lceil\psi\rceil)$ where Bel is a predicate standing for the modal operator B, and $\lceil\alpha\rceil$ is a suitable quotation of the sentence α. However, the difficulties caused by unquoting quoted statements, through substitutions in first-order logic languages have been investigated by Montague in [34] who thus argued in favor of modal logic. In contrast to Montague's thesis, Perlis in [37,38] argues that modal logics are on no firmer ground than first-order logic when equally endowed with substitutive self-reference. Perlis introduces a suitable notion of substitution of a name $\lceil p\rceil$ for its expression p in formulae (providing a similar expressiveness of the gödelian numbering function in the arithmetic language). A function symbol $sub(\lceil P\rceil, \lceil Q\rceil, \lceil\alpha\rceil)$ of substitution is supplied ($sub(\lceil P\rceil, \lceil Q\rceil, \lceil\alpha\rceil) = S$ holds if S is the name for the term obtained as the result of naming the substitution in P of $\lceil Q\rceil$ for all occurrences of $\lceil\alpha\rceil$). Perlis suggests that for an intelligent reasoner such self-referential languages are desirable but, essentially, only limiting results are reported.

In this paper we have shown that in the modal logic $KD4Z$, we can postulate a weaker notion of definability, admitting several fixed points to the self-referential sentence $p \equiv E(p)$, and therefore $KD4Z$ admits a class of modal predicates expressing the provability of consistency. Finally we have addressed how to solve the problem of representing self-reference as an internal construction, thus giving a solution to the problem of representing explicit definability.

Acknowledgements

We thank Dov Gabbay for many useful discussions on fixed points in the logic G and on their computation; Dov has also given important contributions on previous works on provability logic. We thank Ray Reiter for reading a draft of this paper and giving important suggestions on the definability problem, Vladimir Lifshitz for comments on disjunctive definitions, and Jon Doyle for providing us with an unpublished manuscript in which he commented on the relationship between logic G and nonmonotonic reasoning and for a vivid discussion on this point. We thank also the anonymous reviewers for many worthy comments that improved the presentation and for pointing us the earlier investigations on the connection between provability logic and nonmonotonic logic.

Appendix A

A.1. *Basic theorems of KD4Z*

The following are useful theorems of K:

- K1: $\Box(A \rightarrow B) \rightarrow (\Diamond A \rightarrow \Diamond B)$.
- K2: $\Box A \wedge \Diamond B \rightarrow \Diamond(A \wedge B)$.

The following are useful theorems and rules of $KD4Z$:

Let L be $\Diamond\Box$ and M be $\Box\Diamond$.

- T1: $LLp \equiv Lp$ $(MMp \equiv Mp)$.
- T2: $\Box Lp \equiv Lp$, $\Diamond Lp \equiv Lp$ and $L\Box p \equiv Lp$.
- T3: $L(Lp \rightarrow p)$.
- T4: $MLp \rightarrow Lp$.
- T5: $M(q \wedge Mp) \rightarrow Mp$.
- R1: If A is a theorem then LA is a theorem of $KD4Z$. $Lp \rightarrow p$ is not a theorem.

It is easy to show the soundness of the above sentences in $KD4Z$, hence by the completeness theorem they are proved.

A.2. *Self-reference and default logic in KD4Z*

Let $[s]$ be the strong box defined as $[s]A = \Box A \wedge A$.

Proposition A.1 (Formalization lemma (FORL)). *For any A and B the following are equivalent*:

(i) $\vdash_{KD4Z} \Box A \rightarrow \Box B$,

(ii) $\vdash_{KD4Z} [s]A \rightarrow B$.

Proof. One direction follows from transitivity, necessitation and schema K. We give a semantic proof for $\vdash_{KD4Z} \Box\alpha \rightarrow \Box\beta$ implies $\vdash_{KD4Z} [s]\alpha \rightarrow \beta$. From $\vdash_{KD4Z} \Box\alpha \rightarrow \Box\beta$ we have $\vdash_{KD4Z} \Box\alpha \wedge \alpha \rightarrow \Box\beta$, and thus $\vdash_{KD4Z} [s]\alpha \rightarrow \Box\beta$. Suppose that there is a model $\mathfrak{A}$ and a world w, such that $\mathfrak{A}, w \models [s]\alpha$, $w \models \Box\beta$ and $\mathfrak{A}, w \not\models \beta$. Consider the submodel $\mathfrak{B}$ generated by w. $\mathfrak{B}$ is of the form $\mathfrak{F} \odot \mathcal{C}$ with $\mathfrak{F}$ well-capped and with a

first element. Let $\mathfrak{B}'$ be the model obtained from $\mathfrak{B}$ by adding a new *first* world w_0. We have $\mathfrak{B}', w \models [s]\alpha$, hence $\mathfrak{B}', w_0 \models \Box\alpha$ and thus $\mathfrak{B}', w_0 \models \Box\beta$, thus $w \models \beta$, a contradiction. □

Proposition A.2 (First substitution lemma (FSL)). *Let $E(p)$ be a boxed formula.*

$$\vdash_{KD4Z} [s](A \equiv B) \rightarrow (E(A) \equiv E(B)).$$

Proposition A.3 (Second substitution lemma (SSL)). *Let $E(p)$ be a boxed formula.*

$$\vdash_{KD4Z} \Box(A \equiv B) \rightarrow \Box(E(A) \equiv E(B)).$$

The proofs of Propositions A.2 and A.3 can be found in Smorynski [48].

Proposition A.4. *Let $C(p)$ be a formula with p propositional variable. Then*
(1) $\vdash_{K4} \Box C(\top) \rightarrow \Box C(\Box C(\top))$.
(2) $\vdash_{KD4Z} \Diamond \Box C(\top) \rightarrow (\Box C(\top) \equiv \Box C(\Box C(\top)))$.

Proof. Proof of (1).

(i) $\vdash_{KD4Z} \Box C(\top) \rightarrow (\top \equiv \Box C(\top))$ (tautology)
(ii) $\vdash_{K4} \Box\Box C(\top) \rightarrow \Box(\top \equiv \Box C(\top))$ (Nec, K)
(iii) $\vdash_{KD4Z} \Box\Box C(\top) \rightarrow (\Box C(\top) \equiv \Box C(\Box C(\top)))$ (SSL)
(iv) $\vdash_{K4} \Box C(\top) \rightarrow (\Box C(\top) \equiv \Box C(\Box C(\top)))$ (Axiom 4)
(v) $\vdash_{K4} \Box C(\top) \rightarrow \Box C(\Box C(\top))$ (tautology)

Proof of (2).

(vi) $\vdash_{KD4Z} \Box C(\Box C(\top)) \rightarrow (\Box\Box C(\top) \rightarrow \Box C(\top))$ ((3) and tautology)
(vii) $\vdash_{KD4Z} \Box C(\Box C(\top)) \rightarrow \Box(\Box\Box C(\top) \rightarrow \Box C(\top))$ (Nec, K and Axiom 4)
(viii) $\vdash_{KD4Z} \Box C(\Box C(\top)) \rightarrow (\Diamond\Box\Box C(\top) \rightarrow \Box\Box C(\top))$ (Axiom Z)
(ix) $\vdash_{KD4Z} \Box C(\Box C(\top)) \rightarrow (\Diamond\Box C(\top) \rightarrow \Box\Box C(\top))$ ($\vdash_{KD4Z} \Diamond\Box\Box\alpha \equiv \Diamond\Box\alpha$ in $KD4Z$ see Section A.1 (T2))
(x) $\vdash_{KD4Z} \Box C(\top) \rightarrow [s](\top \equiv \Box C(\top))$ ((2), Axiom 4)
(xi) $\vdash_{KD4Z} \Box C(\top) \rightarrow (C(\top) \equiv C\,\Box C(\top))$ (FSL)
(xii) $\vdash_{KD4Z} \Box\Box C(\top) \rightarrow (\Box C(\top) \equiv \Box C(\Box C(\top)))$ (Nec and K)
(xiii) $\vdash_{KD4Z} \Diamond\Box C(\top) \rightarrow (\Box C(\Box C(\top)) \rightarrow \Box C(\top))$ (tautology, (12) and (9))
(xiv) $\vdash_{KD4Z} \Diamond\Box C(\top) \rightarrow (\Box C(\Box C(\top)) \equiv \Box C(\top))$ (tautology, (13) and (5))

□

Theorem A.5 (Definability of predicates of the form $\Box C(p)$). *Let* $E(p) = \Box C(p)$, *with* $E(p)$ *an implicit consistency predicate. Then* $\Box C(\top)$ *is a fixed point. That is*

$$\vdash_{KD4Z} \Box C(\top) \equiv \Box C(\Box C(\top)).$$

Proof.

(i) $\vdash_{KD4Z} \Box C(p) \rightarrow \Box \Diamond p$ (hypothesis)

(ii) $\vdash_{KD4Z} \Box C(\Box C(\top)) \rightarrow \Box \Diamond \Box C(\top)$ (Uniform Substitution)

(iii) $\vdash_{KD4Z} \Box C(\Box C(\top)) \rightarrow \Diamond \Box C(\top)$

($\vdash_{KD4Z} \Box \Diamond p \equiv \Diamond \Box \Diamond p$ see Section A.1 (T2))

(iv) $\vdash_{KD4Z} \Box C(\top) \equiv \Box C(\Box C(\top))$ (Proposition A.4(1) and (2))

□

Proposition A.6. *Let* $E(p) = B(\Box C(p))$ *and suppose* $\vdash_{KD4Z} C(p) \rightarrow \Diamond p$. *Then*:

$$\vdash_{KD4Z} E(B(\top)) \equiv E(E(B(\top))).$$

The proof is as in Smorynski [48], where $C(p)$ is $C(B(p))$.

Proposition A.7. *Let* p *be boxed in* $E(p)$ *and let* q *be a new variable. Then*:

(1) $\vdash_{KD4Z} \Box(p \equiv q) \rightarrow (E(p) \equiv E(q))$.

(2) $\vdash_{KD4Z} [s](E(p) \equiv p) \wedge [s](E(q) \equiv q) \rightarrow (\Box(p \equiv q) \rightarrow (p \equiv q))$.

(3) $\vdash_{KD4Z} [s](E(p) \equiv p) \wedge [s](E(q) \equiv q) \rightarrow (\Diamond \Box (p \equiv q) \rightarrow (p \equiv q))$.

Proof. The proofs of (1) and (2) are as in Smorynski [48].

Proof of (3).

(i) $\vdash_{KD4Z} [s](E(p) \equiv p) \wedge [s](E(q) \equiv q) \rightarrow \Box(p \equiv q) \rightarrow (p \equiv q))$ ((2))

(ii) $\vdash_{KD4Z} \Box(E(p) \equiv p) \wedge \Box(E(q) \equiv q) \rightarrow \Box(\Box(p \equiv q) \rightarrow (p \equiv q))$ (FORL)

(iii) $\vdash_{KD4Z} \Box(E(p) \equiv p) \wedge \Box(E(q) \equiv q) \rightarrow (\Diamond \Box (p \equiv q) \rightarrow \Box(p \equiv q))$ (Z)

(iv) $\vdash_{KD4Z} \Box(E(p) \equiv p) \wedge \Box(E(q) \equiv q) \rightarrow (\Diamond \Box (p \equiv q) \rightarrow (p \equiv q))$ ((i) and tautology)

□

Proposition A.8. $\vdash_{KD4Z} [s](p \equiv \Box C(p)) \rightarrow (p \rightarrow \Box C(\top))$.

Proof.

(i) $\vdash_{KD4Z} \Box p \rightarrow \Box(\top \equiv p)$ (tautology, Nec, K)
(ii) $\vdash_{KD4Z} \Box(\top \equiv p) \rightarrow \Box(C(\top) \equiv C(p))$ (SSL)
(iii) $\vdash_{KD4Z} \Box p \rightarrow \Box(C(\top) \equiv C(p))$ ((i), (ii) and tautology)
(iv) $\vdash_{KD4Z} [s](p \equiv \Box C(p)) \rightarrow ((p \rightarrow \Box p) \equiv (\Box C(p) \rightarrow \Box\Box C(p)))$ (FSL with $E(q) = q \rightarrow \Box q$)
(v) $\vdash_{KD4Z} [s](p \equiv \Box C(p)) \rightarrow (p \rightarrow \Box p)$ ((iv), Axiom 4 and tautology)
(vi) $\vdash_{KD4Z} [s](p \equiv \Box C(p)) \rightarrow (p \rightarrow \Box(C(\top) \equiv C(p)))$ ((v), (iii) and tautology)
(vii) $\vdash_{KD4Z} [s](p \equiv \Box C(p)) \rightarrow (p \rightarrow \Box C(\top))$ ((vi), K and tautology)

□

As a consequence of the above proposition, we can establish:

Theorem A.9 (Existence of the maximum). *Every fixed point $\mathcal{T}$ of $\Box C(p)$ is of the form $\Box C(\top) \wedge \gamma$ for some γ, that is $\mathcal{T} \rightarrow \Box C(\top)$.*

Proposition A.10. *Let $\mathcal{T}$ be a fixed point of $E(p) = B(\Box C_1(p), \ldots, \Box C_n(p))$ and $\vdash_{KD4Z} E(p) \rightarrow \Box \Diamond p$. Then:*

(1) *$\mathcal{T} \wedge \Box \Diamond \alpha$ is a fixed point of $E(p)$;*
(2) *$\mathcal{T} \wedge \Box \Diamond \alpha$ is a consistent fixed point of $E(p)$ iff $\mathcal{T} \nvdash_{KD4Z} \Diamond \Box \neg \alpha$;*
(3) *$\mathcal{T} \wedge \Diamond \Box \alpha$ is a fixed point of $E(p)$;*
(4) *$\mathcal{T} \wedge \Diamond \Box \alpha$ is a consistent fixed point of $E(p)$ iff $\mathcal{T} \nvdash_{KD4Z} \Box \Diamond \neg \alpha$.*

Proof. Proof of (1).

(i) $\vdash_{KD4Z} E(p) \rightarrow \Box \Diamond p$ (hypothesis)
(ii) $\vdash_{KD4Z} E(\mathcal{T}) \equiv \mathcal{T}$ (hypothesis)
(iii) $\vdash_{KD4Z} \Box \Diamond q \rightarrow [s](\mathcal{T} \equiv \mathcal{T} \wedge \Box \Diamond q)$ (q not occurring in $\mathcal{T}$; theorem of $K4$)
(iv) $\vdash_{KD4Z} \Box \Diamond q \rightarrow (E(\mathcal{T}) \equiv E(\mathcal{T} \wedge \Box \Diamond p))$ (FSL)
(v) $\vdash_{KD4Z} \Box \Diamond q \wedge E(\mathcal{T}) \equiv \Box \Diamond q \wedge E(\mathcal{T} \wedge \Box \Diamond q)$ (tautology)
(vi) $\vdash_{KD4Z} \Box \Diamond q \wedge \mathcal{T} \equiv \Box \Diamond q \wedge E(\mathcal{T} \wedge \Box \Diamond q)$ ((ii), tautology)
(vii) $\vdash_{KD4Z} E(\mathcal{T} \wedge \Box \Diamond q) \rightarrow \Box \Diamond (\mathcal{T} \wedge \Box \Diamond q)$ ((ii), Uniform Substitution)
(viii) $\vdash_{KD4Z} \Box \Diamond (\mathcal{T} \wedge \Box \Diamond q) \rightarrow \Box \Diamond q$ (see Section A.1 (T5))
(ix) $\vdash_{KD4Z} \Box \Diamond q \wedge \mathcal{T} \equiv E(\mathcal{T} \wedge \Box \Diamond q)$ ((vi), (viii) and tautology)

Proof of (2). $\mathcal{T} \wedge \Box \Diamond \alpha$ is a consistent fixed point of $E(p)$ iff $\nvdash_{KD4Z} \neg(\mathcal{T} \wedge \Box \Diamond \alpha)$ iff $\nvdash_{KD4Z} \mathcal{T} \rightarrow \Diamond \Box \neg \alpha$ iff $\mathcal{T} \nvdash_{KD4Z} \Diamond \Box \neg \alpha$.

Proof of (3) and (4). Since $\mathcal{T} \wedge \Box \Diamond q$ is a fixed point then substitute $\Box q$ for q. From

$$\vdash_{KD4Z} \Box \Diamond \Box q \equiv \Diamond \Box q$$

we get that $\mathcal{T} \wedge \Diamond \Box q$ is a fixed point. □

Proposition A.11. $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow (p \equiv \Box \Diamond p)$.

Proof.

(i) $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow (\Box p \rightarrow p)$ (Proposition A.8)
(ii) $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow (p \rightarrow \Box p)$ (Proposition A.8(v))
(iii) $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow (p \equiv \Box p)$ ((i) and (ii))
(iv) $\vdash_{KD4Z} \Box p \rightarrow \Diamond \Box p$ (D and Axiom 4)
(v) $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow (p \rightarrow \Diamond p)$ ((ii) and D)
(vi) $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow (\Diamond \Box p \rightarrow \Box p)$
((i), Nec, Axiom 4 and Z)
(vii) $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow [s](p \equiv \Box p)$
((iii), Nec, K and tautology)
(viii) $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow (\Diamond p \rightarrow p)$
((vi), (vii), FSL with $\Diamond p \rightarrow p$)
(ix) $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow (p \equiv \Diamond p)$ ((v), (viii))
(x) $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow (\Box p \equiv \Box \Diamond p)$ (Nec, K)
(xi) $\vdash_{KD4Z} [\underline{s}](p \leftrightarrow \Box C(p)) \wedge \Box C(\top) \rightarrow (p \equiv \Box \Diamond p)$ ((iii))
□

Theorem A.12 (Characterization of the structure of fixed points). *$\mathcal{T}$ is a fixed point of $E(p) = \Box C(p)$, with $E(p)$ an implicit consistent predicate, if and only if $\mathcal{T}$ is logically equivalent to a formula of the form $\Box C(\top) \wedge \Box \Diamond \gamma$. Moreover, the set of all fixed points is*

$\{\Box C(\top) \wedge \Box \Diamond \gamma \mid \gamma$ *any formula*$\}$.

Proof.

(i) $\vdash_{KD4Z} \Box C(\top) \rightarrow (\mathcal{T} \equiv \Box \Diamond \mathcal{T})$ (Proposition A.11 and hypothesis)
(ii) $\vdash_{KD4Z} \mathcal{T} \rightarrow \Box C(\top)$ (Theorem A.9)
(iii) $\vdash_{KD4Z} \mathcal{T} \equiv \Box C(\top) \wedge \Box \Diamond \mathcal{T}$ (tautology)

Furthermore, by Propositions A.11 and A.4 the set Φ of all fixed points of $\Box C(p)$ is the set

$\{\Box C(\top) \wedge \Box \Diamond \gamma \mid \gamma$ any formula$\}$. □

Proposition A.13. *If $\mathcal{T}$ is a saturated fixed point then*

$$\vdash_{\Lambda} \mathcal{T} \equiv Tr_{\langle W,D' \rangle}(\mathcal{T}) \wedge \bigwedge_{\beta \in J_{D'}} \Box \Diamond (\mathcal{T} \wedge \beta)$$

where

$$D' = \left\{ \left\langle \frac{\alpha : \beta}{\gamma} \right\rangle \in D \mid \beta \text{ is consistent with } \mathcal{T}^* \right\}.$$

Proof. Let D' be defined as above. When $\beta \in J_{D'}$, there is a model $\mathfrak{A}$ of $\mathcal{T}$ satisfying $\Diamond\Diamond\beta$ (the models of $S5$ are models of $KD4Z$).

Let us first prove $\vdash_A \mathcal{T} \rightarrow \bigwedge_{\beta \in J_{D'}} \Box\Diamond(\mathcal{T} \wedge \beta)$ in the hypothesis that $\mathcal{T}$ is a saturated fixed point. Then the first half of the equivalence follows from $Tr_{\langle W,D\rangle}(\mathcal{T}) \vdash_{KD4Z} Tr_{\langle W,D'\rangle}(\mathcal{T})$. Let $\mathfrak{A}$ be a model. Since $\vdash_{KD4Z} \mathcal{T} \rightarrow \Box\mathcal{T}$ then, without loss of generality, we may suppose that $\mathfrak{A}$ is a model of $\mathcal{T}$. If $\mathfrak{A}$ satisfies $\Diamond\Diamond\beta$, with $\beta \in J_{D'}$ then, by hypothesis, we have $\mathfrak{A} \models St_{\langle W,D\rangle}(\mathcal{T})$, hence $\mathfrak{A} \models \Box\Diamond\beta$, that is $\mathfrak{A} \models \Box\Diamond(\mathcal{T} \wedge \beta)$.

We may suppose that $\mathfrak{A}$ is a model of $\mathcal{T}$, with an initial world w_0, that does not satisfy $\Diamond\Diamond\beta$ for some $\beta \in J_{D'}$, hence $\mathfrak{A} \models \Diamond\Box\neg\beta$. Let $\mathcal{U}$ be the set of interpretations u of the universal model corresponding to the stable set of $\mathcal{T}^*$. Let us insert $\mathcal{U}$ in the frame of $\mathfrak{A}$ updating the evaluation of all formulae on $\mathcal{U}$ and on the set of worlds preceding $\mathcal{U}$ in the relation R by forcing. We obtain a new model $\mathfrak{A}'$.

$\mathfrak{A}'$ is such that for all modal-free formulae φ $w_0 \models \Diamond\Diamond\varphi$ if φ is consistent with $\mathcal{T}^*$, but $w_0 \not\models \Diamond\Diamond\Diamond\beta$ (this is obtained by inserting $\mathcal{U}$ as a set of siblings of any immediate successor of the initial world w_0). By the hypothesis of consistency of β with $\mathcal{T}^*$, $w_0 \models \Diamond\Diamond\beta$. Obviously the worlds which are successive to $\mathcal{U}$ still satisfy the same set of formulae satisfied in $\mathfrak{A}$, hence the new model $\mathfrak{A}'$ cannot be a model of $\mathcal{T}$, otherwise $\mathfrak{A}' \models \Box\Diamond\beta$, by reasoning as in the first part of the proof, but $\mathfrak{A}' \models \Diamond\Box\neg\beta$. Furthermore $\mathcal{T}$ is equivalent to the boxed formula $Tr_{\langle W,D\rangle}(\mathcal{T})$: since for all worlds w successive to $\mathcal{U}$, $w \models \Diamond\mathcal{T} \bigwedge_{\delta \in D} \wedge(\Box(\mathcal{T} \rightarrow \alpha) \wedge \Box\Diamond(\mathcal{T} \wedge \beta) \rightarrow \Box(\mathcal{T} \rightarrow \gamma)) \wedge \Box W \wedge W$, then $\mathcal{U} \models Tr_{\langle W,D\rangle}(\mathcal{T})$ hence $\mathcal{U} \models \mathcal{T}$. Then for some world w preceding $\mathcal{U}$, $w \not\models \mathcal{T}$. Now from the fact that $\mathcal{T}$ is logically equivalent to $Tr_{\langle W,D\rangle}(\mathcal{T})$, $Tr_{\langle W,D\rangle}(\mathcal{T})$ is not satisfied. This implies that for some default $\langle \frac{\alpha' : \beta'}{\gamma'} \rangle$, with $\beta \neq \beta'$, there is a world w preceding $\mathcal{U}$ and successive to the first world w_0, such that $w \models \Box(\mathcal{T} \rightarrow \alpha') \wedge \Box\Diamond(\mathcal{T} \wedge \beta') \wedge \Diamond(\mathcal{T} \wedge \neg\gamma')$ that is for some world v successive to w we have $v \models \mathcal{T} \rightarrow \alpha'$ and $v \models \mathcal{T} \wedge \neg\gamma'$ and $\beta' \in J_{D'}$. This world v must belong to $\mathcal{U}$, because of the way we have inserted $\mathcal{U}$, hence $\mathcal{U} \models \mathcal{T} \rightarrow \alpha'$ and $v \models \mathcal{T} \wedge \neg\gamma'$, and then, from $\mathcal{U} \models \mathcal{T}$, we get $\mathcal{U} \models \alpha'$ and $v \models \neg\gamma'$, that is $\mathcal{U} \not\models \gamma'$. But from $\vdash_{KD4Z} \mathcal{T} \equiv Tr_{\langle W,D\rangle}(\mathcal{T})$ and $\vdash_{KD4Z} \mathcal{T} \rightarrow \Box\mathcal{T}$ we have $\mathcal{T} \vdash_{KD4Z} \Box\Box\alpha' \wedge \Box\Diamond\beta' \rightarrow \Box\Box\gamma'$. Therefore $\mathcal{T} \vdash_{S5} \Box\alpha' \wedge \Diamond\beta' \rightarrow \Box\gamma'$. Since $\mathcal{U} \models_{S5} \alpha'$ and $\mathcal{U} \models_{S5} \Diamond\beta'$ hence $\mathcal{U} \models_{S5} \gamma'$ which implies $\mathcal{U} \models \gamma'$: we have a contradiction.

As for the other direction of the equivalence, notice that if β is not in D' then $\mathcal{T}$ derives $\Box\Box\neg\beta$ hence $Tr_{\langle W,D'\rangle}(\mathcal{T}) \vdash_A Tr_{\langle W,D\rangle}(\mathcal{T})$. From the hypothesis we have $\vdash_A Tr_{\langle W,D\rangle}(\mathcal{T}) \rightarrow \mathcal{T}$ hence $\vdash_{KD4Z} \bigwedge_{\beta \in J_{D'}} \Box\Diamond(\mathcal{T} \wedge \beta) \wedge Tr_{\langle W,D'\rangle}(\mathcal{T}) \rightarrow \mathcal{T}$. □

Proposition A.14. $\vdash_{KD4Z} (\mathcal{T} \equiv Tr_{\langle W,D\rangle}(\mathcal{T})) \wedge (\mathcal{T} \rightarrow St_{\langle W,D\rangle}(\mathcal{T}))$ *only if* $\mathcal{T}^*$ *is a Reiter extension.*

Proof. $\mathcal{T}^*$ is a Reiter fixed point for the operator Γ when using $\mathcal{T}^*$ as a context. In fact, by Theorem A.9, $\mathcal{T}$ implies $\Box C(\top)$, hence $\Box(\Box\alpha \wedge \Box\Diamond\beta \rightarrow \Box\gamma)$, and by Proposition A.13 considering the reduct D', $\mathcal{T}^*$ is closed with respect to D. On the other hand it is $Cn_{D'}(W) = \mathcal{T}^*$, where $Cn_{D'}(W)$ is the set of all formulae which have a derivation with

defaults D' from the set of axioms W with the context T^*. Let us prove it by induction on the length of a derivation of a formula in $Cn_{D'}(W)$. Assume that $\alpha_1, \ldots, \alpha_n$ is a derivation and for all j, $j \leqslant i$, $\mathcal{T} \vdash_{KD4Z} \Box\Box\alpha_j$: let us prove that $\mathcal{T} \vdash_{KD4Z} \Box\Box\alpha_{i+1}$. The nontrivial case is when α_{j+1} is a consequence γ of a default $\langle \frac{\alpha:\beta}{\gamma} \rangle \in D'$. By inductive hypothesis we have $\mathcal{T} \vdash_{KD4Z} \Box\Box\alpha$ hence, from $\vdash_{KD4Z} \mathcal{T} \rightarrow \Box\Diamond\beta \wedge (\Box\Box\alpha \rightarrow \Box\Box\gamma$, from Proposition A.13, we have $\vdash_{KD4Z} \mathcal{T} \rightarrow \Box\Box\gamma)$, hence $\gamma \in T^*$. By [28, Proposition 3.26] T^* is a Reiter extension of $\langle W, D \rangle$ being D' the reduct of D with respect to T^* and $Cn_{D'}(W) = T^*$. □

References

[1] G. Amati, L. Carlucci Aiello, D. Gabbay and F. Pirri, A structural property on modal frames characterizing default logic, *J. IGPL* **4** (1996) 1–24.

[2] G. Amati, L. Carlucci Aiello and F. Pirri, Deduction methods for KD4Z, Tech. Rept. Rap. VI.94, Universitá di Roma "La Sapienza", Roma (1994).

[3] G. Amati, L. Carlucci Aiello and F. Pirri, Intuitionistic autoepistemic logic, *Stud. Logica* (to appear).

[4] C. Bernardi, The fixed-point theorem for diagonalizable algebras, *Stud. Logica* **34** (1975) 239–251.

[5] P. Besnard and T. Schaub, Possible world semantics for default logic, in: *Working Notes 4th International Workshop on Nonmonotonic Reasoning* (1992) 34–40.

[6] E. Beth, On Padoa's method in the theory of definition, *Indag. Math.* **15** (1953) 330–339.

[7] G. Boolos, *On the Unprovability of Consistency* (Oxford University Press, Oxford, 1979).

[8] J. Doyle, A model for deliberation, action, and introspection, AI-TR 581, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA (1980).

[9] J. Doyle, A mathematical basis for psychology, Unpublished manuscript (1981).

[10] J. Doyle, Some theories of reasoned assumptions: an essay in rational psychology, Tech. Rept. CMU-CS-83-125, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA (1983).

[11] J. Doyle, Circumscription and implicit definability, *J. Automated Reasoning* **1** (1985) 391–405.

[12] D. Etherington, *Reasoning with Incomplete Information* (Morgan Kaufman, San Mateo, CA, 1988).

[13] M. Fitting, *Proof Methods for Modal and Intuitionistic Logics* (Reidel, Dordrecht, 1983).

[14] D.M. Gabbay, Modal provability interpretations for negation by failure, in: P. Shroeder-Heister, ed., *Extensions of Logic Programming*, Lecture Notes in Computer Science **475** (Springer, Berlin, 1991) 179–222.

[15] P. Gärdenfors, A geometric model of concept formation, *Information Modelling and Knowledge Bases* **III** (IOS Press, Amsterdam, 1992) 1–17.

[16] M. Gelfond and V. Lifschitz, The stable model semantics for logic programming, in: *Proceedings Fifth International Conference on Logic Programming (ICLP-88)* (1988) 230–237.

[17] R. Goldblatt, Logics of time and computation, CSLI (1992).

[18] J. Halpern and Y. Moses, Toward a theory of knowledge and ignorance: preliminary report, in: *Proceedings AAAI Workshop on Non-Monotonic Reasoning*, New Paltz, NY (Morgan Kaufmann, Los Altos, CA, 1984) 125–143.

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

[20] H.J. Levesque, All I know: a study in autoepistemic logic, *Artif. Intell.* **42** (1990) 381–386.

[21] V. Lifschitz, Computing circumscription, in: *Proceedings IJCAI-85*, Los Angeles, CA (1985) 121–127.

[22] V. Lifschitz, Formal theories of action, in: *The Frame Problem in Artificial Intelligence* (Morgan Kaufmann, Los Altos, CA, 1987) 35–57.

[23] F. Lin, Specifying the effects of indeterminate actions, Tech. Rept., Computer Science Department, University of Toronto, Toronto, Ont. (1996).

[24] F. Lin and R. Reiter, State constraints revisited, *J. Logic Comput.* **4** (1994) 655–678.

[25] F. Lin and Y. Shoham, Epistemic semantics for fixed-points non-monotonic logics, in: *Proceedings Third Conference on Theoretical Aspects of Reasoning about Knowledge*, Pacific Grove, CA (1990) 111–120.

[26] V.W. Marek, G. Schwarz and M. Truszczynski, Modal nonmonotonic logics: ranges, characterization, computation, *J. ACM* **40** (1993) 963–990.
[27] V.W. Marek and M. Truszczynski, Modal logic for default reasoning, in: *Ann. Math. Artif. Intell.* **1** (1990) 275–302.
[28] V.W. Marek and M. Truszczynski, *Nonmonotonic Logic, Context-Dependent Reasoning* (Springer, Berlin, 1993).
[29] J. McCarthy, Circumscription—a form of nonmonotonic reasoning, *Artif. Intell.* **13** (1980) 27–39.
[30] D. McDermott, Non-monotonic logic II: non-monotonic modal theories, *J. ACM* **29** (1982) 33–57.
[31] D. McDermott and J. Doyle, Non-monotonic logic I, *Artif. Intell.* **13** (1980) 41–72.
[32] C. Mervis and E. Rosh, Categorization of natural objects, *Annual Rev. Psychol.* **32** (1981) 89–115.
[33] F. Montagna, The predicate modal logic of provability, *Notre Dame J. Formal Logic* **25** (1984) 179–189.
[34] R. Montague, Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability, *Acta Philos. Fennica* **16** (1963) 153–167.
[35] R. Moore, Semantical considerations on nonmonotonic logics, *Artif. Intell.* **25** (1985) 75–94.
[36] R. Moore, Possible-world semantics for autoepistemic logic, in: M. Ginsberg, ed., *Readings in Nonmonotonic Reasoning* (Morgan Kaufmann, Los Altos, CA, 1987) 137–142.
[37] D. Perlis, Languages with self-reference I: foundations, *Artif. Intell.* **25** (1985) 301–322.
[38] D. Perlis, Languages with self-reference II: knowledge, belief, and modality, *Artif. Intell.* **34** (1988) 179–212.
[39] R. Reiter, A logic for default reasoning, *Artif. Intell.* **13** (1980) 81–132.
[40] R. Reiter, Circumscription implies predicate completion (sometimes), in: *Proceedings AAAI-82*, Pittsburgh, PA (1982) 183–188.
[41] R. Reiter, Nonmonotonic reasoning, in: *Annual Review of Computer Science* **2** (Annual Reviews Inc., Palo Alto, CA, 1987) 147–186.
[42] G. Schwarz, Autoepistemic modal logics, in: *Proceedings Third Conference on Theoretical Aspects of Reasoning about Knowledge*, Pacific Grove, CA (1990) 97–109.
[43] G. Schwarz, Minimal model semantics for nonmonotonic modal logics, in: *Proceedings First International Workshop on Logic Programming and Nonmonotonic Reasoning*, Washington, DC (1991) 260–274.
[44] G. Schwarz, Minimal model semantics for nonmonotonic modal logics, in: *Proceedings Seventh Annual IEEE Symposium on Logic in Computer Science*, Santa Cruz, CA (1992) 34–43.
[45] G. Schwarz, In search of a "true" logic of knowledge: the nonmonotonic perspective, *Artif. Intell.* **79** (1995) 39–63.
[46] G. Schwarz and M. Truszczynski, Modal logic S4f and the minimal knowledge paradigm, in: *Proceedings Third Conference on Theoretical Aspects of Reasoning about Knowledge*, Pacific Grove, CA (1990) 97–109.
[47] G. Schwarz and M. Truszczynski, Minimal knowledge problem: a new approach, *Artif. Intell.* **67** (1994) 113–141.
[48] C. Smorynski, *Self-Reference and Modal Logic* (Springer, Berlin, 1985).
[49] C. Smorynski, Quantified modal logic and self-reference, *Notre Dame J. Formal Logic* **28** (1987) 356–370.
[50] R. Smullyan, Languages in which self-reference is possible, *J. Symbolic Logic* **22** (1957) 55–67.
[51] R. Smullyan, *Diagonalization and Self-Reference* (Oxford University Press, Oxford, 1994).
[52] R. Solovay, Provability interpretations of modal logic, *Israel J. Math.* **25** (1976) 287–304.
[53] R. Stalnaker, A note on non-monotonic modal logic, *Artif. Intell.* **64** (1993) 183–196.
[54] A. Tarski, *Some Methodological Investigations on the Definability of Concepts* (Oxford University Press, Oxford, 1956).
[55] M. Truszczynski, Modal interpretations of default logic, in: *Proceedings IJCAI-91*, Sydney (1991) 393–398.