Circumscription and Definability

Yves Moinard
IRISA, Campus de Beaulieu
35042 RENNES-Cedex, FRANCE
tel: (33) 99-36-20 00
E-mail: moinard@irisa.fr

Raymond Rolland
IRMAR, Campus de Beaulieu
35042 RENNES-Cedex, FRANCE.
tel: (33) 99-28-60-19

Abstract

Thanks to two stronger versions of predicate circumscription (one of the best known non-monotonic reasoning methods), we give a definitive answer to two old open problems. The first one is the problem of expressing domain circumscription in terms of predicate circumscription. The second one is the problem of definability of the circumscribed predicates, asked by Doyle in 1985, and never answered since. These two results, and the way used to obtain them, could help an "automatic circumscriptor".

1 Introduction

Firstly, McCarthy defined domain circumscription which reduces the set of individuals ($2). Later, he defined predicate circumscription, which reduces the extensions of some relations (§3). McCarthy has stated [I980| that domain circumscription is a particular case of predicate circumscription. [Etherington and Mercer, 1987] reaffirmed the importance of domain circumscription and contested McCarthy's statement. We show (58) why this contestation is not fully justified and we provide two improvements of McCarthy's translation. To obtain our results, we define (§4) a variant of the strong pointwise circumscription of [Lifschitz, 1988a]. Cases of equivalence with standard circumscrition (§5) allow us to answer the central question in [Doyle, 1985]: when does circumscription define the predicates (§6)? A stronger circumscription, "definabilization" (§7), simplifies the expression, and hopefully the computation, of domain circumscription. We make precise the expression of domain circumscription in terms of predicate circumscription. Throughout the text we provide the semantics for each kind of "circumscription" defined, thus all of them, including the first order versions, may be considered as preferential entailment notions.

2 Domain circumscription

In many situations only the objects named are supposed to exist. Domain circumscription ([McCarthy, 1980], amended by [Morreau, 1985, Etherington and Mercer, 1987]) formalizes this idea. A theory T is a set of formulas in a first order language C, $ or *[xo), is a formula in $\mathcal{L}$, $\Phi[x]$ denotes the formula obtained by substituting the free variable x for every free occurrence of x_0 in Φ (x not bound in $\Phi[x_0]$).

Definitions 2.1 a) $Axiom_{\mathcal{L}}(\Phi)$ denotes $\{\exists x_0\ \Phi[x_0]\}$, together with the set of formulas $\{\forall \mathbf{x}\ (\widehat{\Phi[\mathbf{x}]} \Rightarrow \Phi[f(\mathbf{x})])/f$ is a function symbol in $\mathcal{L}\}$. $\mathbf{x}$ is a tuple $(x_1, ..., x_k)$ of free variables, k is the arity of f, and $\widehat{\Phi[\mathbf{x}]}$ is $\Phi[x_1] \wedge ... \wedge \Phi[x_k]$ ($TRUE$ if $k=0$).

b) $\mathcal{T}^{\Phi}$, the *relativization of $\mathcal{T}$ to Φ*, is $\mathcal{T}$ in prenex form, where for each variable symbol y, we replace $\forall y\ (\cdots)$ by $\forall y\ (\Phi[y] \Rightarrow (\cdots))$ and $\exists y\ (\cdots)$ by $\exists y\ (\Phi[y] \wedge (\cdots))$ (thus $P(a)$ by $\Phi[a] \wedge P(a)$).

c) Circd($\mathcal{T}$), the *domain circumscription of $\mathcal{T}$*, adds the following axiom schema to $\mathcal{T}$:

(SDC) $(Axiom_{\mathcal{L}}(\Phi) \wedge \mathcal{T}^{\Phi}) \Rightarrow \forall x_0\ \Phi[x_0]$,
for each formula Φ in $\mathcal{L}$.

Notations Our models are normal (equality interpreted as identity). If μ is an interpretation over $\mathcal{L}$, D_μ is its domain. For each predicate P (resp. function f) of arity k in $\mathcal{L}$, the subset $|P|_\mu$ of D_μ^k (resp. the application f_μ of D_μ^k in D_μ) denotes its extension in μ.

$\mathcal{L}_\mu$, the *language of* μ, adds to $\mathcal{L}$ a name for each element in D_μ.

Definition 2.2 (see e.g. [Enderton, 1972] p.90–91) μ and ν are interpretations over $\mathcal{L}$. We denote $\nu < \mu$, if and only if: 1) $D_\nu \subset D_\mu$ (strict inclusion),
2) for each function f in $\mathcal{L}$, if $e \in D_\nu^k$, then $f_\mu(e) = f_\nu(e)$,
3) for each predicate P in $\mathcal{L}$, $|P|_\mu \cap D_\nu^k = |P|_\nu$.

Definition 2.3 $\mathcal{R}$ is a binary relation between interpretations over $\mathcal{L}$. A model μ of $\mathcal{T}$ is *minimal for* $\mathcal{R}$ when no model ν of $\mathcal{T}$ is such that $\nu \mathcal{R} \mu$.

Theorem 2.1 (Davis [1980]) Any model of $\mathcal{T}$ minimal for $<$ is a model of **Circd($\mathcal{T}$)**.

The converse needs a more sophisticated relation:

Definition 2.4 (see e.g.[Enderton, 1972] p.88–89 and ex. 24 p.97): A subset S of D_μ^k is *definable with parameters in* μ when there exists a formula Φ in $\mathcal{L}_\mu$, of arity k, such that $S = |\Phi|_\mu$.

Definition 2.5 If $\nu < \mu$ and if D_ν is definable with parameters in μ, we note $\nu <^\delta \mu$.

Theorem 2.2 ([Morreau, 1985], or §8) The models of **Circd($\mathcal{T}$)** are the models of $\mathcal{T}$ minimal for $<^\delta$.

Here is an example showing how domain circumscription reduces the domain to the minimum allowed by the axioms of the given theory (note that here, it does not matter whether we use $<$ or $<^\delta$).

Example 2.1 $\mathcal{T}$ is $P(a) \wedge \exists x \neg P(x)$.
$Axiom_{\mathcal{L}}(\Phi)$ is $\Phi[a]$, $\mathcal{T}^{\Phi}$ is $\Phi[a] \wedge P(a) \wedge \exists x(\Phi[x] \wedge \neg P(x))$ (SDC) is: $\mathcal{T}^{\Phi} \Rightarrow \forall x\ \Phi[x]$ (for every Φ in $\mathcal{L}$). Choosing $(x = a \vee x = y)$ as $\Phi[x]$ in (SDC) gives: $\mathbf{Circd}(\mathcal{T}) \equiv P(a) \wedge \exists y(\neg P(y) \wedge \forall x\ (x = a \vee x = y))$. The models of $\mathbf{Circd}(\mathcal{T})$ are the models of $\mathcal{T}$ with only two elements.

3 Predicate circumscription

Definition 3.1 ([McCarthy, 1986]) **a)** Let $\mathcal{T}$ be a finite set of first order formulas in which the lists of predicates $\mathbf{P} = (P_1, \cdots, P_n)$ and $\mathbf{Q} = (Q_1, \cdots, Q_m)$ occur. The *first order circumscription* of $\mathbf{P}$ in $\mathcal{T}$ with $\mathbf{Q}$ varying, noted $\mathbf{Circ}_1(\mathcal{T}\colon \mathbf{P};\ \mathbf{Q})$ ($\mathbf{Circ}_1(\mathcal{T}\colon \mathbf{P})$ if $\mathbf{Q}$ is empty), adds the following axiom schema to $\mathcal{T}$, (SAC):
$\{\mathcal{T}[\mathbf{p},\mathbf{q}] \wedge \forall \mathbf{x}\ (\mathbf{p}[\mathbf{x}] \Rightarrow \mathbf{P}(\mathbf{x}))\} \Rightarrow \forall \mathbf{x}\ (\mathbf{P}(\mathbf{x}) \Rightarrow \mathbf{p}[\mathbf{x}])$,
for every $\mathbf{p} = (p_1, \cdots, p_n)$, $\mathbf{q} = (q_1, \cdots, q_m)$ lists of first order formulas in $\mathcal{L}$.

$\mathcal{T}[\mathbf{p},\mathbf{q}]$ is $\mathcal{T}$ where each occurrence of P_i (resp. Q_j) is replaced by p_i (resp. q_j). $\mathbf{p}[\mathbf{x}] \Rightarrow \mathbf{P}(\mathbf{x})$ is $(p_1[\mathbf{x}_1] \Rightarrow P_1(\mathbf{x}_1)) \wedge \cdots \wedge (p_n[\mathbf{x}_n] \Rightarrow P_n(\mathbf{x}_n))$.

The square brackets mean that $p_i[x_i]$ may have free variables not included in $\mathbf{x}_i = \{x_{i,1}, \cdots, x_{i,k_i}\}$ (k_i arity of P_i) (see [Besnard *et al.*, 1989] and ex. 8.1 below).

b) The *second order circumscription* $\mathrm{Circ}_2(\mathcal{T}\colon \mathbf{P};\ \mathbf{Q})$ adds the following axiom to $\mathcal{T}$, (AC):
$\forall \mathbf{pq}\ \{[\mathcal{T}[\mathbf{p},\mathbf{q}] \wedge (\forall \mathbf{x}\ (\mathbf{p}(\mathbf{x}) \Rightarrow \mathbf{P}(\mathbf{x}))] \Rightarrow \forall \mathbf{x}\ (\mathbf{P}(\mathbf{x}) \Rightarrow \mathbf{p}(\mathbf{x}))\}$,
$\mathbf{p}$ and $\mathbf{q}$ are lists of predicate variables, each p_i (resp. q_j) having the arity of P_i (resp. Q_j).

Circ denotes either $\mathbf{Circ}_1$ or $\mathbf{Circ}_2$.

Definition 3.2 Let μ and ν be two interpretations over $\mathcal{L}$. We write $\mu =_{\mathbf{P};\mathbf{Q}} \nu$ when μ and ν are identical except that there is no condition on the extensions of the P_j's $(1 \le j \le n)$ and Q_j's $(1 \le j \le m)$. Furthermore, if each $|P_i|_\mu$ and $|Q_j|_\mu$ is definable with parameters in ν, we write $\mu =^{\delta}_{\mathbf{P};\mathbf{Q}} \nu$. If Q is empty, we write $=_{\mathbf{P}}$ or $=^{\delta}_{\mathbf{P}}$.

$=_{\mathbf{P};\mathbf{Q}}$ is an equivalence relation, $=^{\delta}_{\mathbf{P}}$ is reflexive and transitive but it is not symmetrical (ex. 5.3).

Definition 3.3 $\mu <_{\mathbf{P};\mathbf{Q}} \nu$ (respectively $\mu <^{\delta}_{\mathbf{P};\mathbf{Q}} \nu$) means $\mu =_{\mathbf{P};\mathbf{Q}} \nu$ (respectively $\mu =^{\delta}_{\mathbf{P};\mathbf{Q}} \nu$), and $|P_i|_\mu \subseteq |P_i|_\nu$ $(1 \le i \le n)$ with some $|P_i|_\mu \subset |P_i|_\nu$.

If $\mathbf{Q} = \emptyset$, we note $<_{\mathbf{P}}$ or $<^{\delta}_{\mathbf{P}}$.

Definition 3.4 (Lifschitz [1986]) $\mathcal{T}$ is *(*$\mathbf{P}$*; Q)-well founded* when for every model μ of $\mathcal{T}$, not minimal for $<_{\mathbf{P};\mathbf{Q}}$, there exists a model ν of $\mathcal{T}$ minimal for $<_{\mathbf{P};\mathbf{Q}}$ with $\nu <_{\mathbf{P};\mathbf{Q}} \mu$.

Universal theories are ($\mathbf{P}$; $\mathbf{Q}$)-well founded [Bossu and Siegel, 1985, Etherington *et al.*, 1985, Lifschitz, 1986].

Theorem 3.1 a) *Soundness* (see [McCarthy, 1980, Lifschitz, 1986]): Every model of $\mathcal{T}$ which is minimal for $<_{\mathbf{P};\mathbf{Q}}$ is a model of $\mathbf{Circ}(\mathcal{T}\colon \mathbf{P};\ \mathbf{Q})$.

b) [Lifschitz, 1986] The models of $\mathbf{Circ}_2(\mathcal{T}\colon \mathbf{P};\ \mathbf{Q})$ are the models of $\mathcal{T}$ minimal for $<_{\mathbf{P};\mathbf{Q}}$.

c) [Besnard, 1989] The models of $\mathbf{Circ}_1(\mathcal{T}\colon \mathbf{P};\ \mathbf{Q})$ are the models of $\mathcal{T}$ minimal for $<^{\delta}_{\mathbf{P};\mathbf{Q}}$.

A model minimal for $<_{\mathbf{P};\mathbf{Q}}$ is minimal for $<^{\delta}_{\mathbf{P};\mathbf{Q}}$.

As it is one of the purposes of this text, we give now an example where a seemingly stronger axiom is in fact equivalent to the circumscription axiom (first order or second order version).

Example 3.1 $\mathcal{T}$ is $P(0) \wedge \forall x\ (P(x) \Rightarrow P(s(x)))$.

$\mathbf{Circ}_1(\mathcal{T} : P)$, adds the following axiom schema to $\mathcal{T}$: for each formula p in $\mathcal{L}$: (SAC) $\{p[0] \wedge \forall x\ (p[x] \Rightarrow p[s(x)])\} \wedge \forall x\ (p[x] \Rightarrow P(x))\} \Rightarrow \forall x\ (P(x) \Rightarrow p[x])$. Let (SAR) be the axiom schema of (Peano's) recurrence: for each p in $\mathcal{L}$: $\{p[0] \wedge \forall x\ (p[x] \Rightarrow p[s(x)])\} \Rightarrow \forall x\ (P(x) \Rightarrow p[x])$. Clearly (SAR) entails (SAC), so $\mathcal{T} \cup$ (SAR) entails $\mathbf{Circ}_1(\mathcal{T} : P)$. Here the converse is true (§5).

4 A strong predicate circumscription

We reinforce and simplify (SAC) the circumscription schema, or the axiom (AC).

Definition 4.1 The first order strong circumscription of $\mathbf{P}$ in $\mathcal{T}$, with $\mathbf{Q}$ varying, $\mathbf{Circf}_1(\mathcal{T}\colon \mathbf{P};\ \mathbf{Q})$, adds to $\mathcal{T}$ the schema (SACf): $\mathcal{T}[\mathbf{p},\mathbf{q}] \Rightarrow \forall \mathbf{x}\ (\mathbf{P}(\mathbf{x}) \Rightarrow \mathbf{p}[\mathbf{x}])$. The 2^{nd} order version $\mathbf{Circf}_2(\mathcal{T}\colon \mathbf{P};\ \mathbf{Q})$, adds (ACf) $\equiv \forall \mathbf{pq}\ \{\mathcal{T}[\mathbf{p},\mathbf{q}] \Rightarrow \forall \mathbf{x}\ (\mathbf{P}(\mathbf{x}) \Rightarrow \mathbf{p}(\mathbf{x}))\}$ to $\mathcal{T}$.

Circf denotes either $\mathbf{Circf}_1$ or $\mathbf{Circf}_2$.

$\mathbf{Circf}(\mathcal{T} : \mathbf{P}; \mathbf{Q}) \vdash \mathbf{Circ}(\mathcal{T} : \mathbf{P}; \mathbf{Q})$.

Here is the semantics for this strong circumscription.

Definitions 4.2 We write $\nu \prec_{\mathbf{P};\mathbf{Q}} \mu$ (respectively $\nu \prec^{\delta}_{\mathbf{P};\mathbf{Q}} \mu$) when $\nu =_{\mathbf{P};\mathbf{Q}} \mu$ (respectively $\nu =^{\delta}_{\mathbf{P};\mathbf{Q}} \mu$) and $|P_i|_\mu - |P_i|_\nu \neq \emptyset$ for at least one i $(1 \le i \le n)$.

"$-$" denotes the set difference.

If $\mathbf{Q}$ is empty, then we write $\prec_{\mathbf{P}}$ or $\prec^{\delta}_{\mathbf{P}}$.

Theorem 4.1 The models of $\mathbf{Circf}_2(\mathcal{T}\colon \mathbf{P};\ \mathbf{Q})$ (respectively $\mathbf{Circf}_1(\mathcal{T}\colon \mathbf{P};\ \mathbf{Q})$) are the models of $\mathcal{T}$ minimal for $\prec_{\mathbf{P};\mathbf{Q}}$ (respectively $\prec^{\delta}_{\mathbf{P};\mathbf{Q}}$).

Remarks 4.1 a) The proofs are as the proofs of theorem 3.1.

b) If $\nu <_{\mathbf{P};\mathbf{Q}} \mu$, then $\nu \prec_{\mathbf{P};\mathbf{Q}} \mu$.

c) $\prec_{\mathbf{P};\mathbf{Q}}$ is neither transitive nor antisymmetrical.

d) $(\nu \prec_{P_1,P_2;\mathbf{Q}} \mu)$ if and only if
$(\nu \prec_{P_1;P_2,\mathbf{Q}} \mu)$ or $(\nu \prec_{P_2;P_1,\mathbf{Q}} \mu)$.

b), c) and d) also hold with $^{\delta}$-relations.

Example 4.1 $\mathcal{T}$ is $(P(a) \vee P(b)) \wedge a \neq b$.

— $\mathbf{Circ}(\mathcal{T} : P)$ is consistent, being equivalent to
$a \neq b \wedge (\forall x\ (P(x) \Leftrightarrow x = a) \vee \forall x\ (P(x) \Leftrightarrow x = b))$.
— From $\mathbf{Circf}(\mathcal{T} : P)$, $\neg P(a)$ can be proved (choose $x = b$ as $p[x]$ in (SACf)), as can $\neg P(b)$ (choose $p[x] \equiv x = a$). Thus $\mathbf{Circf}(\mathcal{T} : P)$ is inconsistent.
— Here are 3 models of $\mathcal{T}$, with the same domain

$\{a,b,e\}$: $|P|_\mu = \{a,b,e\}$, $|P|_\nu = \{a\}$, $|P|_{\nu'} = \{b\}$. ν and ν', minimal for $<_P$, are models of **Circ**($\mathcal{T}$: P): $\nu <_P \mu$, $\nu' <_P \mu$. $\nu \prec_P \mu$, $\nu' \prec_P \mu$, $\nu \not\prec_P \nu'$ and $\nu' \not\prec_P \nu$: there is no model minimal for $\prec_P$. Here, the domain is finite and so the $^\delta$-relations coincide with these relations.

Here is a result about strong circumscription (proof: remark 4.1-d, which is obvious):

Theorem 4.2 **Circf**($\mathcal{T}$: $(P_1, ..., P_n); \mathbf{Q}$) is equivalent to $\bigcup_{j=1}^{n}$ **Circf**($\mathcal{T}$: $P_j; P_1, ..., P_{j-1}, P_{j+1}, ..., P_n, \mathbf{Q}$).

Now, this strong circumscription is not as new as it may seem, although the presentation and the semantics are new:

Definition 4.3 ([Lifschitz, 1988b]) The *strong pointwise circumscription* adds to $\mathcal{T}$ the schema (SAPf): $\forall \mathbf{x}\ \neg(P(x) \wedge \neg p[x] \wedge \mathcal{T}[p, \mathbf{q}])$, for every tuple of formulas $(p, \mathbf{q})$ in $\mathcal{L}$.

We have: (SAPf) $\equiv \forall \mathbf{x}\ (\mathcal{T}[p, \mathbf{q}] \Rightarrow (P(\mathbf{x}) \Rightarrow p[\mathbf{x}])) \equiv$
$(\mathcal{T}[p, \mathbf{q}] \Rightarrow \forall \mathbf{x}\ (P(\mathbf{x}) \Rightarrow p[\mathbf{x}])) \equiv$ (SACf).

Lifschitz defines only *individual* strong pointwise circumscription (for one P at a time), and uses unions. We allow tuples **P** in *parallel* strong circumscription. Theorem 4.2 shows that these two ways are equivalent. As this circumscription is not pointwise at all, we prefer our name "strong circumscription".

Lifschitz gives an elegant second order formula

$$\mathbf{Circf}_2(\mathcal{T} : \mathbf{P}; \mathbf{Q}) \equiv \mathcal{T} \wedge \forall \mathbf{x}\ \{\mathbf{P}(\mathbf{x}) \Leftrightarrow (\mathrm{L})\}, \text{ with}$$
$$(\mathrm{L}) \equiv \forall \mathbf{pq}\ (\mathcal{T}[\mathbf{p}, \mathbf{q}] \Rightarrow \mathbf{p}(\mathbf{x})).$$

Also, the *"modal strong circumscription"* of [Perlis, 1988], differs from (SACf) only in the fact that $\mathbf{P}(\mathbf{x})$ is replaced by $K\mathbf{P}(\mathbf{x})$, where K is a modal operator, attenuating this otherwise too strong circumscription.

5 When strong and standard circumscriptions coincide

Definitions 5.1 **a)** A theory $\mathcal{T} \equiv \mathcal{T}[\mathbf{P}; \mathbf{Q}]$ is *stable for extended conjunction in* (**P**; **Q**) (first order version) if, for every tuple of formulas $\mathbf{p}, \mathbf{q}, \mathbf{p}', \mathbf{q}'$ in $\mathcal{L}$:

$\mathcal{T}[\mathbf{p}, \mathbf{q}] \cup \mathcal{T}[\mathbf{p}', \mathbf{q}'] \vdash \mathcal{T}[\mathbf{p} \wedge \mathbf{p}' \wedge \mathbf{p}'', \mathbf{q}'']$ for some tuples $\mathbf{p}'', \mathbf{q}''$ of formulas in $\mathcal{L}$.

$\mathbf{p}, \mathbf{p}', \mathbf{p}''$ have the same length as the tuple of predicates **P**, and $\mathbf{q}, \mathbf{q}', \mathbf{q}''$ have the length of **Q**. $\mathbf{p} \wedge \mathbf{p}' \wedge \mathbf{p}''$ denotes the tuple of the $p_i \wedge p_i' \wedge p_i''$'s involved.

b) $\mathcal{T}$ is *stable for extended conjunction in* (**P**; **Q**) (second order version) if $\forall \mathbf{p}\,\mathbf{q}\,\mathbf{p}'\,\mathbf{q}'\ \exists \mathbf{p}''\,\mathbf{q}''\ \{(\mathcal{T}[\mathbf{p}, \mathbf{q}] \wedge \mathcal{T}[\mathbf{p}', \mathbf{q}']) \Rightarrow \mathcal{T}[\mathbf{p} \wedge \mathbf{p}' \wedge \mathbf{p}'', \mathbf{q}'']\}$ is true in $\mathcal{T}$. $\mathbf{p}, \mathbf{q}, \mathbf{p}', \mathbf{q}', \mathbf{p}'', \mathbf{q}''$ are tuples of predicate variables of suitable lengths and arities.

If there is no $\mathbf{p}''$, $\mathcal{T}$ is *stable for conjunction in* **P**.

Theorem 5.1 If $\mathcal{T}$ is stable for extended conjunction in (**P**; **Q**), then **Circf**($\mathcal{T}$: **P**; **Q**) $\equiv$ **Circ**($\mathcal{T}$: **P**; **Q**).

<u>Proof</u>: This theorem is true for the 1^{st} and for the 2^{nd} order versions of definitions 3.1, 4.1 and 5.1. We give the 1^{st} order version.

Circf$_1$($\mathcal{T}$: **P**; **Q**) entails **Circ**$_1$($\mathcal{T}$: **P**; **Q**), we need the converse. **Circ**$_1$($\mathcal{T}$: **P**; **Q**) is: $\mathcal{T}$ and (SAC). Suppose we also have $\mathcal{T}[\mathbf{p},\mathbf{q}]$. $\mathcal{T} \equiv \mathcal{T}[\mathbf{P}, \mathbf{Q}]$, so we get (definition 5.1) $\mathcal{T}[\mathbf{p} \wedge \mathbf{P} \wedge \mathbf{p}'', \mathbf{q}'']$, for some $\mathbf{p}''$, $\mathbf{q}''$. The instance of (SAC) associated with $(\mathbf{p} \wedge \mathbf{P} \wedge \mathbf{p}'', \mathbf{q}'')$ gives $\forall \mathbf{x}\ (\mathbf{P}(\mathbf{x}) \Rightarrow (\mathbf{p}[\mathbf{x}] \wedge \mathbf{P}(\mathbf{x}) \wedge \mathbf{p}''[\mathbf{x}]))$. Thus $\mathcal{T}[\mathbf{p}, \mathbf{q}] \Rightarrow \forall \mathbf{x}\ (\mathbf{P}(\mathbf{x}) \Rightarrow \mathbf{p}[\mathbf{x}])$, i.e. (SACf)

Horn theories are stable for conjunction in (**P**; **Q**) (generalize a well-known result of [van Emden and Kowalski, 1976]). Thus, for Horn theories (or for stratified logic programs, see [Moinard, 1990]) we may simplify the circumscription axiom. So in example 3.1, **Circf**$_1$($\mathcal{T}$: P), which is $\mathcal{T} \cup$ (SAR) in this case, is equivalent to **Circ**$_1$($\mathcal{T}$: P).

Definition 5.2 A theory $\mathcal{T}$ has the property of *extended intersection in* (**P**; **Q**) if and only if, for every models μ and ν of $\mathcal{T}$ such that $\nu =_{\mathbf{P};\mathbf{Q}} \mu$, there exists a model μ' of $\mathcal{T}$ such that $\mu' =_{\mathbf{P};\mathbf{Q}} \mu$, with $|\mathbf{P}|_{\mu'} \subseteq |\mathbf{P}|_\mu \cap |\mathbf{P}|_\nu$.

Theorem 5.2 **a)** A theory $\mathcal{T}$ is stable for extended conjunction in (**P**; **Q**) (second order version), if and only if $\mathcal{T}$ has the property of extended intersection in (**P**; **Q**).

b) If $\mathcal{T}$ is (**P**; **Q**)-well founded, then **Circ**$_2$($\mathcal{T}$: **P**; **Q**) is equivalent to **Circf**$_2$($\mathcal{T}$: **P**; **Q**) if and only if $\mathcal{T}$ has the property of extended intersection in (**P**; **Q**).

<u>Proof</u>: **a)** Details lengthy, but obvious (see [Moinard and Rolland, 1991].

b) Theorem 5.1 gives one way. **Circ**$_2$($\mathcal{T}$: **P**; **Q**) $\equiv$ **Circf**$_2$($\mathcal{T}$: **P**; **Q**) if and only if: for every model μ' of $\mathcal{T}$ minimal for $<_{\mathbf{P};\mathbf{Q}}$, if μ is a model of $\mathcal{T}$ and $\mu =_{\mathbf{P};\mathbf{Q}} \mu'$, we have: $|\mathbf{P}|'_\mu \subseteq |\mathbf{P}|_\mu$ (see remark 4.1 b). Let μ and ν be two models of $\mathcal{T}$ with $\nu =_{\mathbf{P};\mathbf{Q}} \mu$. As $\mathcal{T}$ is (**P**; **Q**)-well founded, there exists μ' model of $\mathcal{T}$, minimal for $<_{\mathbf{P};\mathbf{Q}}$ with ($\mu' <_{\mathbf{P};\mathbf{Q}} \mu$ or $\mu' = \mu$). $\mu' =_{\mathbf{P};\mathbf{Q}} \mu$, so $|\mathbf{P}|_{\mu'} \subseteq |\mathbf{P}|_\mu$. Also, $\mu' =_{\mathbf{P};\mathbf{Q}} \nu$ thus $|\mathbf{P}|_{\mu'} \subseteq |\mathbf{P}|_\nu$ and $|\mathbf{P}|_{\mu'} \subseteq |\mathbf{P}|_\mu \cap |\mathbf{P}|_\nu$.

If $\mathcal{T}$ is (**P**; **Q**)-well founded, we may refine this result:

Definition 5.3 $\mathcal{T} \equiv \mathcal{T}[\mathbf{P}; \mathbf{Q}]$ has the property of *unique minimal model in* (**P**; **Q**) if and only if, for every models μ and ν of $\mathcal{T}$ minimal for $<_{\mathbf{P};\mathbf{Q}}$ and such that $\mu =_{\mathbf{P};\mathbf{Q}} \nu$, we have $|\mathbf{P}|_\mu = |\mathbf{P}|_\nu$ (i.e. $\mu =_{\mathbf{Q}} \nu$).

Theorem 5.3 **a)** If $\mathcal{T}$ has the property of extended intersection in (**P**; **Q**), then $\mathcal{T}$ has the property of unique minimal model in (**P**; **Q**).

b) If $\mathcal{T}$ is (**P**; **Q**)-well founded, then $\mathcal{T}$ has the property of extended intersection in (**P**; **Q**) if and only if $\mathcal{T}$ has the property of unique minimal model in (**P**; **Q**).

<u>Proof</u>: **a)** Obvious.

b) Let μ and ν be models of $\mathcal{T}$ with $\mu =_{\mathbf{P};\mathbf{Q}} \nu$. There exist models μ' and ν' of $\mathcal{T}$ minimal for $<_{\mathbf{P};\mathbf{Q}}$ with ($\mu' =_{\mathbf{P};\mathbf{Q}} \mu$, $|\mathbf{P}|_{\mu'} \subseteq |\mathbf{P}|_\mu$) and ($\nu' =_{\mathbf{P};\mathbf{Q}} \nu$, $|\mathbf{P}|_{\nu'} \subseteq |\mathbf{P}|_\nu$). Thus $\mu' =_{\mathbf{P};\mathbf{Q}} \nu'$, and $|\mathbf{P}|_{\mu'} = |\mathbf{P}|_{\nu'}$.

Example 5.1 $\mathcal{T}$: $P(a)$, $P(b) \Rightarrow (P(c) \vee P(d))$,
$\{a \neq b, b \neq c, b \neq d, c \neq d, a \neq c, a \neq d\}$ {(S1)}.

— **Circf**($\mathcal{T}$: P) $\equiv$ **Circ**($\mathcal{T}$: P) $\equiv$
(S1) $\wedge(\forall x\ (P(x) \Leftrightarrow x = a)$ (choose $p(x) \equiv x = a$ in AC).
— $\mathcal{T}$ is not stable for conjunction in P, but $\mathcal{T}$ is stable for extended conjunction in P ($p''(x) \equiv x = a$). $\mathcal{T}$ is

P-well founded, and has the property of unique minimal model in P. This example shows that we must use the *extended* conjunction in theorem 5.2.

Example 5.2 $\mathcal{T}$ is:

$P(0) \vee \exists x \, \{P(x) \wedge \forall y \, (P(y) \Rightarrow (P(s(y)) \wedge x \neq s(y)))\}$,
$\forall x \forall y \, (s(x) = s(y) \Rightarrow x = y), \forall x \, (s(x) \neq 0)$ (S2)

— $\mathcal{T}$ is not P-well founded (right-hand side of disjunction).

— $\mathbf{Circ}_2(\mathcal{T} : P) \equiv$ (S2) $\wedge \forall x \, (P(x) \Leftrightarrow x = 0)$ (consistent).

$\mathbf{Circf}_2(\mathcal{T} : P)$ entails $P(0)$ (choose $x = 0$ as $p(x)$ in (SACF)), and also $\neg P(0)$ ($p(x) \equiv x \neq 0$), thus it is inconsistent.

— $\mathcal{T}$ has the property of unique minimal model in P: for every model μ of $\mathcal{T}$ minimal for $<_P$, $|P|_\mu = \{0_\mu\}$. $\mathcal{T}$ does not have the property of extended intersection in P: take $D_\mu = \mathbb{N} = \{0, 1, 2, 3, \ldots\}$, s_μ successor function in $\mathbb{N}$, $|P|_\mu = \{0\}$, $0_\mu = 0$; $\nu =_P \mu$, $|P|_\nu = \{1, 2, 3, \ldots\}$. μ and ν are two models of $\mathcal{T}$, $|P|_\mu \cap |P|_\nu = \emptyset$, and as $\mathcal{T}$[FALSE] is not true, no model μ' of $\mathcal{T}$ verifies $\mu' =_P \mu$ and $|P|_{\mu'} \subseteq |P|_\mu \cap |P|_\nu$. μ is minimal for $<_P$, ν is not (choose $|P|_{\nu'} = \{2, 3, 4, \ldots\}$). $\nu \prec_P \mu$, μ is not minimal for $\prec_P$ (no model is minimal for $\prec_P$). This example shows that we need the well-foundedness condition in theorem 5.3-b.

Other examples [Moinard and Rolland, 1991] show the importance of the well foundedness condition also for theorem 5.2 b or show that, since theorem 5.2, the main results are false with the first order versions (even with an adapted notion of well-foundedness). Thus, in the end of this section and in the following section, we deal only with the second order versions. Here we give only a simple example of one problem encountered with first order versions:

Example 5.3 $\mathcal{T}$: $P(0) \wedge \forall x \, (P(x) \Rightarrow P(s(x)))$.

(cf example 3.1). It is easy to find two models of $\mathbf{Circf}_1(\mathcal{T} : P)$ with $\mu =_P \nu$ and $|P|_\mu \neq |P|_\nu$. Choose $\mathbb{N} \cup \mathbb{Z}$ for domain, $|P|_\mu = \mathbb{N}$, $|P|_\nu = \mathbb{N} \cup \mathbb{Z}$. The problem is that $\mu =^\delta_P \nu$ is false ($|P|_\mu$ is not definable with parameters in ν) thus $\mu <^\delta_P \nu$ is false.

(Note that $\nu =^\delta_P \mu$ is true: $|P|_\nu = D_\mu$ is definable in μ).

6 Definability of the circumscribed predicate

Here are two definitions reminded by [Doyle, 1985]. This definability ***in a theory*** must not be confused with the definability ***in a model*** of definition 2.4.

Definition 6.1 A first order theory $\mathcal{T}$ ***implicitly defines*** P if and only if whenever μ and ν are two models of $\mathcal{T}$ such that $\nu =_P \mu$, then they are identical.

Definition 6.2 A first order theory $\mathcal{T}$ ***explicitly defines*** P if and only if there exists a first order formula Φ in $\mathcal{L}$, not involving P, such that: $\mathcal{T} \vdash \forall x \, (P(x) \Leftrightarrow \Phi)$.

The ***Beth's definability theorem*** (see [Chang and Keisler, 1973, Doyle, 1985]), states that implicit definability is equivalent to explicit definability. Now, we give a definition adapted to circumscription with variable predicates:

Definition 6.3 A 2^{nd} ***order circumscription*** of $\mathbf{P}$ in $\mathcal{T}$ with $\mathbf{Q}$ varying ***implicitly defines*** $\mathbf{P}$ if and only if $\mathcal{T}$ has the property of unique minimal model in $(\mathbf{P}; \mathbf{Q})$.

This is a natural extension of definition 6.1, when "varying predicates" are involved. As already mentioned, we must use the second order circumscription, but $\mathcal{T}$ is a first order theory, and we may consider only first order models. If $\mathbf{Q}$ is empty, definition 6.3 coincides with definition 6.1: $\mathbf{Circ}_2(\mathcal{T} : P)$ implicitly defines P (definition 6.3) if and only if whenever μ and ν are models of $\mathbf{Circ}_2(\mathcal{T} : P)$ such that $\nu =_P \mu$, then $\mu = \nu$ (cf definition 6.1). So, the theorems 5.2 and 5.3 give:

Theorem 6.1 a) If $\mathcal{T}$ is such that $\mathbf{Circ}_2(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$ is equivalent to $\mathbf{Circf}_2(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$, then $\mathbf{Circ}_2(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$ implicitly defines $\mathbf{P}$.

b) If $\mathcal{T}$ has the property of extended conjunction in $(\mathbf{P}; \mathbf{Q})$ (second order version),then $\mathbf{Circ}_2(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$ implicitly defines $\mathbf{P}$.

c) If $\mathcal{T}$ is $(\mathbf{P}; \mathbf{Q})$-well founded, then $\mathbf{Circ}_2(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$ implicitly defines $\mathbf{P}$ if and only if $\mathbf{Circ}_2(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$ is equivalent to $\mathbf{Circf}_2(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$, that is if and only if $\mathcal{T}$ is stable for extended conjunction in $(\mathbf{P}; \mathbf{Q})$ (second order version).

This gives the answer to the central problem in [Doyle, 1985], asking when P is definable in the circumscribed theory. Also it strengthens the importance of formula (L), given at the end of §4, with standard circumscription. (L) does not involve $\mathbf{P}$ nor $\mathbf{Q}$, so for well-founded theories, (L) gives the explicit second order definition of $\mathbf{P}$ in $\mathbf{Circ}_2(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$ in all the cases where such a definition does exist. Note however that (L) being a second order formula is **not** the Φ of definition 6.2; such a Φ does not necessarily exist because $\mathbf{Circ}_2(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$ is a second order theory (cf example below).

Examples In examples 5.1, 5.2 and 5.3, $\mathbf{Circ}_2(\mathcal{T} : P)$ implicitly defines P. We may even give an explicit first order definition (for first or second order circumscription) in examples 5.1 ($P(x) \equiv x = a$) and 5.2 ($P(x) \equiv x = 0$). In example 5.3, $\mathbf{Circ}_1(\mathcal{T} : P)$ does not implicitly defines P. Also it is a case where no <u>first order</u> explicit definition exists for P in $\mathbf{Circ}_2(\mathcal{T} : P)$.

7 Definabilization (a stronger circumscription)

Definition 7.1 $\mathbf{Def}_1(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$, the ***first order definabilization of*** $\mathbf{P}$ in $\mathcal{T}$ with $\mathbf{Q}$ varying, adds to $\mathcal{T}$ the axiom schema (SADf): $\mathcal{T}[\mathbf{p}, \mathbf{q}] \Rightarrow \forall \mathbf{x} \, (\mathbf{P}(\mathbf{x}) \Leftrightarrow \mathbf{p}[\mathbf{x}])$. The 2^{nd} order version $\mathbf{Def}_2(\mathcal{T}\text{: } \mathbf{P}; \mathbf{Q})$ adds to $\mathcal{T}$ (ADf) $\equiv \forall \mathbf{p}, \mathbf{q} \, \{\mathcal{T}[\mathbf{p}, \mathbf{q}] \Rightarrow \forall \mathbf{x} \, (\mathbf{P}(\mathbf{x}) \Leftrightarrow \mathbf{p}(\mathbf{x}))\}$.

Def denotes $\mathbf{Def}_1$ or $\mathbf{Def}_2$.

$\mathbf{Def}(\mathcal{T} : \mathbf{P}; \mathbf{Q}) \vdash \mathbf{Circf}(\mathcal{T} : \mathbf{P}; \mathbf{Q})$.

Doyle [1985] uses a similar notion without variable predicate, and calls it "implicit definability" (cf definition 6.1). A specific name makes a clear difference between a theory having this property and the addition of an axiom schema to any theory.

Example 7.1 $\mathcal{T}$: $P(a)$ (Horn theory).

— $\mathbf{Circf}(\mathcal{T}:P) \equiv \mathbf{Circ}(\mathcal{T}:P) \equiv \forall x\ (P(x) \Leftrightarrow x=a)$.
— $\mathbf{Def}(\mathcal{T}:P) \equiv P(a) \wedge \forall x\ (x=a)$
(Choose $p[x] \equiv x=a$, then $p[x] \equiv x=x$, in (SADf)).

Example 7.2 $\mathcal{T}$: $P(a)$, $\exists x\ \neg P(x)$ (cf example 2.1).

— $\mathbf{Circf}(\mathcal{T}:P) \equiv \mathbf{Circ}(\mathcal{T}:P) \equiv \forall x\ (P(x) \Leftrightarrow x=a) \wedge \exists x\ \neg P(x)$. ($\mathcal{T}$ is stable for conjunction in P, and $|P|_\mu = \{a_\mu\}$ in every model μ minimal for $<_P$).
— $\mathbf{Def}(\mathcal{T}:P) \equiv$
$\forall x\ (P(x) \equiv x=a) \wedge \exists y\ (y \neq a \wedge \forall x\ (x=a \vee x=y))$.
(Choose $p[x] \equiv x=a$, getting $\forall x\ (P(x) \equiv x=a)$, then $p[x] \equiv x \neq y$ which gives $y \neq a \Rightarrow \forall x\ (P(x) \Leftrightarrow x \neq y)$).
So definabilization minimizes the extensions of the predicates and the domain, as precised now:

Theorem 7.1 a) If $\mathcal{T} \equiv \mathcal{T}[\mathbf{P},\mathbf{Q}]$ entails $\mathcal{T}[\mathbf{TRUE},\mathbf{q}]$ for some tuple of formulas $\mathbf{q}$, then $\mathbf{Def}(\mathcal{T}: \mathbf{P};\mathbf{Q})$ entails $\forall \mathbf{x}\ \mathbf{P}(\mathbf{x})$. ($\mathbf{TRUE}$ is the tuple $(\mathbf{x}_1=\mathbf{x}_1,\dots,\mathbf{x}_n=\mathbf{x}_n)$).
b) $\mathbf{Def}(\mathcal{T}:\mathbf{P};\mathbf{Q}) \vdash \mathbf{Circ}(\mathcal{T}:\mathbf{P};\mathbf{Q})$.
$\{\forall \mathbf{x}\ \mathbf{P}(\mathbf{x})\} \cup \mathbf{Circ}(\mathcal{T}:\mathbf{P};\mathbf{Q}) \vdash \mathbf{Def}(\mathcal{T}:\mathbf{P};\mathbf{Q})$.
c) If $\mathcal{T} \vdash \mathcal{T}[\mathbf{TRUE},\mathbf{q}]$, then
$\mathbf{Def}(\mathcal{T}:\mathbf{P};\mathbf{Q}) \equiv \{\forall \mathbf{x}\ \mathbf{P}(\mathbf{x})\} \cup \mathbf{Circ}(\mathcal{T}:\mathbf{P};\mathbf{Q})$.

Proof: **a)** Choose (**TRUE**, q) as (p,q) in (SADf) or (ADf) (depending of the version).
b) — Obvious.
— $\{\forall \mathbf{x}\ \mathbf{P}(\mathbf{x})\} \cup \mathbf{Circ}(\mathcal{T}:\mathbf{P};\mathbf{Q})$ is: $\mathcal{T} \cup \{\forall \mathbf{x}\ \mathbf{P}(\mathbf{x})\} \cup$ (SAC). $\forall \mathbf{x}\ \mathbf{P}(\mathbf{x})$ gives $\forall x\ (\mathbf{p}[\mathbf{x}] \Rightarrow \mathbf{P}(\mathbf{x}))$, thus (SAC) gives (SADf). Adaptation easy for the 2^{nd} order version.

Definitions 7.2 We write $\nu \#_{\mathbf{P};\mathbf{Q}} \mu$ (resp. $\nu \#^{\delta}_{\mathbf{P};\mathbf{Q}} \mu$) when $\nu =_{\mathbf{P};\mathbf{Q}} \mu$ (resp. $\nu =^{\delta}_{\mathbf{P};\mathbf{Q}} \mu$) and $|P_i|_\mu \neq |P_i|_\nu$ for at least one i $(1 \leq i \leq n)$.

Theorem 7.2 The models of $\mathbf{Def}_2(\mathcal{T}: \mathbf{P};\mathbf{Q})$ (respectively $\mathbf{Def}_1(\mathcal{T}: \mathbf{P};\mathbf{Q})$) are the models of $\mathcal{T}$ minimal for $\#_{\mathbf{P};\mathbf{Q}}$ (respectively $\#^{\delta}_{\mathbf{P};\mathbf{Q}}$).

Proofs obvious. As Doyle [1985] has noted, definabilization is generally too strong, here is an example:

Example 7.3 $\mathcal{T}$: $P(a) \Rightarrow P(b)$, $a \neq b$ (Horn theory).
— $\mathbf{Circf}(\mathcal{T}:P) \equiv \mathbf{Circ}(\mathcal{T}:P) \equiv a \neq b \wedge \forall x\ \neg P(x)$
— $\mathbf{Def}(\mathcal{T}:P)$ is inconsistent:
choosing $p[x] \equiv x=b$ proves $\neg P(a)$, then
choosing $p[x] \equiv (x=a \vee x=b)$ proves $P(a)$.
— Here are 3 models of $\mathcal{T}$, with the same domain $\{a,b\}$: $|P|_\mu = \{b\}, |P|_\nu = \{a,b\}, |P|_{\mu'} = \emptyset$.
$\mu' <_P \mu <_P \nu$. μ', minimal for $<_P$ and $\prec_P$, is a model of $\mathbf{Circ}(\mathcal{T}:P)$ and $\mathbf{Circf}(\mathcal{T}:P)$. $\mu' \#_P \mu$, $\mu \#_P \mu'$, etc... There is no model minimal for $\#_P$.

8 From domain to predicate circumscription

We will use only the first order versions (cf §2). Here is the way McCarthy [1980] uses to transform a domain circumscription into a predicate circumscription:

Definition 8.1 U being a new unary predicate symbol, we note: $\mathcal{T}_U \equiv Axiom_{\mathcal{L}}(U) \wedge \mathcal{T}^U$ and
$\mathcal{T}\langle U\rangle \equiv \mathbf{Circ}_1(\mathcal{T}_U : U) \cup \{\forall x\ U(x)\}$.
$\mathcal{L}(U)$ denotes $\mathcal{L}$ augmented by U. $Axiom_{\mathcal{L}}(U)$ is also $Axiom_{\mathcal{L}(U)}(U)$, as $\mathcal{L}$ and $\mathcal{L}(U)$ have the same function symbols, and we will often note $Axiom(U)$.

Theorem 8.1 (McCarthy [1980]) $\mathcal{T}\langle U\rangle$ simulates the domain circumscription of $\mathcal{T}$: $\mathcal{T}\langle U\rangle$ entails $\mathbf{Circd}(\mathcal{T})$. The converse is true: every formula Φ in $\mathcal{L}$ entailed by $\mathcal{T}\langle U\rangle$, is entailed by $\mathbf{Circd}(\mathcal{T})$. That is, $\mathcal{T}\langle U\rangle$ is a *conservative extension* of $\mathbf{Circd}(\mathcal{T})$.

In a contestation of this method, Etherington and Mercer [1987] argue that consistency is not guaranteed when adding $\forall x\ U(x)$. However, no example is given. We prove that in fact, if an inconsistency has to appear, then it is detected in the circumscription of U:

Theorem 8.2 If $\mathbf{Circ}_1(\mathcal{T}_U : U)$ is consistent, then $\mathcal{T}\langle U\rangle$ is consistent.

Proof: Let μ be a model of $\mathbf{Circ}_1(\mathcal{T}_U : U)$, that is a model of $\mathcal{T}_U$, minimal for $<^{\delta}_P$. We define μ_U (*μ restricted to U*): $\mu_U < \mu$; $D_{\mu_U} = |U|_{\mu_U} = |U|_\mu$. μ_U is a model of $\forall x\ U(x)$. μ is a model of $\mathcal{T}_U$, and so is μ_U. Let us suppose that there exists ν, model of $\mathcal{T}_U$, with $\nu <^{\delta}_U \mu_U$. Let ϕ be a formula in $\mathcal{L}_{\mu_U}$ with: $|\phi|_{\mu_U} = |U|_\nu$ (see definitions 3.3, 2.4). We define μ': $\mu' <_U \mu$, $|U|_{\mu'} = |\phi|_{\mu_U}$. Let Φ be a prenex formula in $\mathcal{L}_{\mu_U}$, we may prove by induction on the length of Φ that: $|\Phi|_{\mu_U} = |\Phi^U|_\mu$ [Moinard and Rolland, 1991]. First, this proves that μ' is a model of $\mathcal{T}^U$. Also, ν is a model of $Axiom_{\mathcal{L}}(U)$, and so is μ', thus μ' is a model of $\mathcal{T}_U$. From: $|U|_{\mu'} = |\phi|_{\mu_U} = |\phi^U|_\mu$, we get $\mu' <^{\delta}_U \mu$ contradicting the fact that μ is a model of $\mathcal{T}_U$ minimal for $<^{\delta}_U$. So there exists no ν and μ_U is a model of $\mathcal{T}_U$ minimal for $<^{\delta}_U$, i.e. a model of $\mathbf{Circ}_1(\mathcal{T}_U : U)$. If there exists a model (μ) of $\mathbf{Circ}_1(\mathcal{T}_U : U)$, there exists a model ($\mu_U$) of $\mathcal{T}\langle U\rangle$. (Proof simpler for the 2^{nd} order version).

Thus, McCarthy's way of expressing domain circumscription into predicate circumscription is justified. One of the main problems in any circumscription is the possibility of unexpected inconsistency. Here, if inconsistency arises it is detected in the circumscribing process, not in the addition of the last axiom $\forall x\ U(x)$ which is harmless.

However, there is a little problem remaining. Any known method of circumscription adds axioms to the initial theory. Here, we leave the initial theory for a while (neither $\mathcal{T}_U$ nor $\mathbf{Circ}_1(\mathcal{T}_U : U)$ entails $\mathcal{T}$), and at the end, we recover $\mathcal{T}$. This does not simplify the matter if we are to automatize the process: for this purpose, we do not want to prove any axiom of $\mathcal{T}$, we need to know that these formulas are true. One theoretical solution meeting this requirement is to introduce $\mathcal{T}$ in the circumscription involved. That is why we propose to circumscribe U in $\mathcal{T} \wedge \mathcal{T}_U$, instead of $\mathcal{T}_U$ alone. As U does not occur in $\mathcal{T}$, we get: $\mathbf{Circ}_1(\mathcal{T}_U : U) \wedge \mathcal{T} \equiv \mathbf{Circ}_1(\mathcal{T}_U \wedge \mathcal{T} : U)$. As $\mathcal{T}\langle U\rangle$ entails $\mathcal{T}$, we have:

Theorem 8.3 $\mathcal{T}\langle U\rangle \equiv \mathbf{Circ}_1(\mathcal{T}_U \wedge \mathcal{T} : U) \cup \{\forall x\ U(x)\}$.
Again, inconsistency cannot be provoqued by the addition of $\forall x\ U(x)$.
If $\mathcal{T}$ is universal, we get a simplification:
$\mathcal{T}\langle U\rangle \equiv \mathbf{Circ}_1(Axiom(U) \wedge \mathcal{T} : U) \cup \{\forall x\ U(x)\}$.
(Indeed, $\mathcal{T}$ entails $\mathcal{T}^U$).

With this simulation of domain circumscription, we are closer to the other known kinds of circumscription: we add an axiom schema to $\mathcal{T}$. "More work is done by the circumscription of U" (the final addition of $\forall x\ U(x)$ eliminates fewer models). As with McCarthy's method, if we had an automatic demonstrator including a predicate circumscriptor (we may tend towards this goal), then we would also have a domain circumscriptor. We cannot avoid adding $\forall x\ \ U(x)$ if we want to use a predicate circumscription (see [Etherington *et al.*, 1985, Etherington and Mercer, 1987]). But now we show how, using definabilization, we may greatly simplify the process and avoid the need for $\forall x\ U(x)$.

Theorem 8.4 $\mathcal{T}<U> \equiv \mathbf{Circ}_1(\mathcal{T}_U \wedge \mathcal{T} : U) \cup \{\forall x\ U(x)\}$
$\equiv \mathbf{Def}_1(\mathcal{T}_U \wedge \mathcal{T} : U)$.

Proof: $\mathcal{T}_U$[TRUE] is $\mathcal{T}^U$[TRUE] $\wedge$ Axiom(TRUE), i.e. $\mathcal{T} \wedge$ Axiom(TRUE), i.e. $\mathcal{T}$, and $\mathcal{T}$[TRUE] is $\mathcal{T}$ because U does not occur in $\mathcal{T}$. Thus: $\mathcal{T}_U[U] \wedge \mathcal{T}[U]$ entails $\mathcal{T}_U$[TRUE] $\wedge$ $\mathcal{T}$[TRUE]. Use theorems 7.1 and 8.1.

Example 8.1 $\mathcal{T}$: $P(a) \wedge \exists x\ \neg P(x)$ (cf ex. 2.1, 7.2).

$Axiom(U) \equiv U(a)$; $\mathcal{T}^U \equiv \mathcal{T}_U \equiv U(a) \wedge \exists x\ (U(x) \wedge \neg P(x))$. $\mathbf{Circ}_1(\mathcal{T}_U : U) \equiv \mathcal{T}_U \cup \{ [u[a] \wedge \exists x\ (u[x] \wedge \neg P(x)) \wedge \forall x\ (u[x] \Rightarrow U(x))] \Rightarrow \forall x\ (U(x) \Rightarrow u[x])$, for any formula u in $\mathcal{L}(U)$ }. We choose $u[x] \equiv (x = a \vee x = y)$, which gives: $(U(y) \wedge \neg P(y)) \Rightarrow \forall x\ (U(x) \Rightarrow (x = a \vee x = y))$. Adding $\forall x\ U(x)$, we get $\neg P(y) \Rightarrow \forall x\ (x = a \vee x = y)$, which together with $P(a) \wedge \exists x\ \neg P(x)$ gives: $P(a) \wedge \exists y\ (\neg P(y) \wedge \forall x\ (x = a \vee x = y)$, that is $\mathbf{Circd}(\mathcal{T})$ (cf example 2.1).

$\mathbf{Def}_1(\mathcal{T}_U : U) \equiv \mathcal{T}_U \cup \{[u[a] \wedge \exists x\ (u[x] \wedge \neg P(x))] \Rightarrow \forall x\ (U(x) \Leftrightarrow u[x])$, for every formula u in $\mathcal{L}(U)$ }. We choose $u[x] \equiv (x = a \vee x = y)$, which gives: $\neg P(y) \Rightarrow \forall x\ (U(x) \Leftrightarrow (x = a \vee x = y))$, then we choose $u[x] \equiv x = x$, which gives: $\forall x\ (x = a \vee x = y)$).

Theorem 8.5 If $\mathcal{T}$ is universal, then we get a simplification: $\mathcal{T}<U> \equiv \mathbf{Def}_1(\mathcal{T} \wedge Axiom(U) : U) \equiv \mathcal{T} \wedge \mathbf{Def}_1(\text{Axiom}(U) : U)$ (U does not appear in $\mathcal{T}$).

The axiom schema of definabilization is simpler than the axiom schema of domain circumscription, so this is an application of definabilization.

9 Conclusion

We h ave precised the definitions, semantics, and possible uses, of two kinds of "super circumscriptions". We have given new cases where the circumscription schema may be simplified, a result which is of theoretical and practical importance, as it could be of some help in the process of automatization of circumscription. These results have solved an old question: when does circumscription uniquely define the circumscribed predicates? Our answer is complete for well founded theories. At last, we have precised and justified the passage from domain circumscription to predicate circumscription. We have shown that this passage is safe: it cannot bring inconsistancy. Also we have given two new methods. The first one enhances the role of predicate circumscription, which is useful if we want to use an automatic predicate circumscriptor for domain circumscription. The second one greatly simplifies the schemas involved.

Acknowledgement We are glad to thank Philippe Besnard who initiated this work.

References

[Besnard *et* a/., 1989] P. Besnard, R. Mercer, and Y. Moinard. The importance of open and recursive circumscription. *Artificial Intel.,* 39:251-262, 1989.

[Besnard, 1989] Philippe Besnard, An *Introduction to Default Logic.* Springer Verlag, 1989.

[Bossu and Siegel, 1985] G. Bossu and P. Siegel. Saturation, nonmonotonic reasoning and the closed-world assumption. *Artificial Intelligence,* 25:13-63, 1985.

[Chang and Keisler, 1973] C.C, Chang and HJ. Keisler. *Model Theory.* North-Holland, 1973.

[Davis, 1980] M. Davis. The mathematics of non-monotonic reasoning. *Artificial Intel.,* 13:73-80,1980.

[Doyle, 1985] Jon Doyle. Circumscription and implicit definability. *Automated Reasoning,* 1:391-405, 1985.

[Enderton, 1972] Herbert B. Enderton. *A Mathematical introduction to logic.* Academic Press, 1972.

[Etherington and Mercer, 1987] D.W. Etherington and R. Mercer. Domain circumscription: a reevaluation. *Computational Intelligence,* 3:94-99, 1987,

[Etherington *et al,* 1985] D.W. Etherington, R. M ercer, and R. Reiter. On the adequacy of predicate circumscription for closed-world reasoning. *Computational Intelligence,* 1:11-15, 1985.

[Lifschitz, 1986] V. Lifschitz. On the satisfiability of circumscription. *Artificial Intelligence,* 28:17-27, 1986.

[Lifschitz, 1988a] V. Lifschitz. On the declarative semantics of logic programs with negation. In J. Minker, ed., *Foundations of Deductive Databases and Logic Programs,* pp. 177-192. Morgan-Kaufmann, 1988.

[Lifschitz, 1988b] V. Lifschitz. Pointwise circumscription. *Readings in Nonmonotonic Reasoning,* Ginsberg ed., pp. 179-193, Morgan-Kaufmann, 1988.

[McCarthy, 1980] John McCarthy. Circumscription-a form of non-monotonic reasoning. *Arttfictal Intelligence,* 13:27-39, 1980.

[McCarthy, 1986] John McCarthy. Application of circumscription to formalizing common sense knowledge. *Artificial Lntelligence,* 28:89-116, 1986.

[Moinard and Holland, 1991] Circumscription and definability (2). T.R, IRISA, Rennes, 1991.

[Moinard 1990] Y. Moinard. Circumscription and Horn theories. In *ECAI,* pages 449-454, 1990.

[Morreau, 1985] M.P. Morreau. Circumscription: A sound and complete form of non-monotonic reasoning. Technical Report 15, University, Amsterdam, 1985.

[Perlis, 1988] Donald Perlis. Autocircumscription. *Artificial Intelligence,* 36:223-236, 1988.

[van Emden and Kowalski, 1976] M,H. van Emden and R.A, Kowalski. The semantics of predicate logic as a programming language. *JACM,* 23(4):841-862, 1976.