# Value Minimization in Circumscription

Chitta Baral, Alfredo Gabaldon [1]

*Dept. of Computer Science, University of Texas at El Paso, El Paso, TX 79968*

Alessandro Provetti [2]

*C.I.R.F.I.D., Università di Bologna, Via Galliera 3, I-40121 Bologna, Italy*

**Abstract**

We introduce and motivate the notion of *value minimizing* a function in circumscription. Value minimization concerns limiting the interpretation of functions/terms to those elements of the universe which are minimal w.r.t. a *weight* criterion, understood as the cost we incur in choosing them. We show how Lifschitz's Nested Abnormality Theories make it possible to axiomatize both function minimization and the weight criteria within a theory which is easy to understand and modify.

Ever since Circumscription was defined, it has been possible to vary function symbols in the minimization process, i.e. to select minimal models irrespective of the interpretation of the function symbol. The best known example is perhaps Baker's theory of action effects where function *Result* is let to vary while the effects of actions are minimized. On the other hand, the idea of minimizing a term is not found –to the best of our knowledge– anywhere in the circumscription literature.

Lifschitz's Nested Abnormality Theories allow the minimization of several predicates, each w.r.t. a subset of the theory. This is done as follows: an abnormality predicate is minimized w.r.t. a subtheory, understood as defining a predicate or a function to which some default assumption apply.

Nested Abnormality Theories (NATs) seem to us the best methodology proposed so far for classical-logic axiomatizations of commonsense reasoning. NATs allow us to single out and discuss the issue of *minimizing term values*, i.e. to make models of a theory interpret a term into an element of the

[1] {*chitta,alfredo*}*@cs.utep.edu*

[2] *provetti@cirfid.unibo.it*

[3] Group Web page: *http://www.cs.utep.edu/csdept/krgroup.html*

domain which is minimal w.r.t. an externally-defined criterion. This criterion is understood as the *cost* which we incur in choosing such an object. When the domain is explicit, i.e. when each individual thereof has a term to represent itself in the theory, then the term minimization criterion can be specified by a NAT, thus achieving compact representations of knowledge.

In this note we extend Lifschitz's work and make the case for applying circumscription to terms as a convenient and sound technique for commonsense reasoning. In Section 3 we introduce an explicit definition of value minimization. We continue in the following section with a model-theoretic definition, similar to Lifschitz's for standard predicate circumscription. Finally, we show in Section 5 how value minimization is captured using predicate circumscription within the framework of NATs.

## 1 Two examples of treating functions

Before discussing value minimization in general, let us illustrate how varying and minimizing functions has been used very recently in theories of actions and change. Both examples concern a NAT axiomatization.

Starting from Baker's [4] treatment of action in circumscription, Kartha and Lifschitz in [4] formalize actions with non-deterministic effects and non-frame fluents. One important step in defining the theory is to axiomatize the notion that "the result of performing an action is normally defined." This is done by postulating the axiom

$$Result(a,s) = \perp \;\supset\; Ab(a,s)$$

and minimizing $Ab$, which implies that a state –denoted *Result(a,s)*– is set to indefinite only when other axioms of the theory imply it to be indefinite.

Recently in [4], we have discussed the relationship between specifications in the action description language $\mathcal{L}$ [4] and NAT-style circumscriptive theories. There we were faced with the minimization of a particular term, $Sit\_map(S_N)$, that mapped the current-situation symbol $S_N$ onto a sequence of actions, understood as the history of the domain, e.g. a model of the theory would be such that:

$\mathcal{M} \models Sit\_map(S_0) = \epsilon$

...

$\mathcal{M} \models Sit\_map(S_N) = A_k \circ A_{k-1} \circ \ldots \epsilon$

where terms $A_k \circ \ldots A_2 \circ A_1 \circ \epsilon$ are understood as "$A_1$ then $A_2$ then ...etc." Minimizing the length of the sequence of actions assigned to $S_N$ was required to formalize the assumption *"no actions occurred except those needed to explain the facts in the theory"* which is present in the semantics of $\mathcal{L}$.

**Example 1** *(discovering the occurrence of actions)*

*Consider the following simple story: initially, $F$ is known to be false; at a later moment $F$ is observed to be true; it is also known that $A$ is the only action in the domain that causes $F$ to become true. This story is described in $\mathcal{L}$ as follows:*

$A$ **causes** $F$

$\neg F$ **at** $S_0$

$F$ **at** $S_1$

$S_0$ **precedes** $S_1$

*Intuitively, we would like to conclude that action $A$ occurred in the initial situation causing $F$ to become true, and no other action has occured since or before. Taking the translation of this theory into NATs, it is easy to check that there is only one model $\mathcal{M}$ of the translation such that:*

$\mathcal{M} \models Sit\_map(S_0) = \epsilon$

$\mathcal{M} \models Sit\_map(S_1) = A \circ \epsilon$

$\mathcal{M} \models Sit\_map(S_N) = A \circ \epsilon$

To achieve the result shown above, the key point is including in each translation $T_{\mathcal{L}}$ of a $\mathcal{L}$-theory the following axiom [4]:

$$Subsequence(\alpha, Sit\_map(S_N)) \supset Ab(\alpha)$$

with the obvious intention to minimize the extent of $Ab$. Indeed, suppose an interpretation $\mathcal{I}$ satisfies all the axioms of $T_{\mathcal{L}}$, and maps $Sit\_map(S_N)$ on the sequence of actions $\alpha$, while interpretation $\mathcal{I}'$ maps $Sit\_map(S_N)$ —other things being equal— onto a supersequence $\beta$ of $\alpha$. As a result, the extent of $Ab$ under $\mathcal{I}$ is a proper subset of the extent of $Ab$ under $\mathcal{I}'$. Therefore, $\mathcal{I}'$ is not a circumscriptive model of $T_{\mathcal{L}}$.

[4] It can be interesting to see the NAT defining predicate *Subsequence* in Appendix A. Please refer to [4] for a complete discussion of the approach.

## 2 Explicit domains

An axiomatic theory makes a domain explicit (Reiter [4] terms it *domain closure assumption*) when for each element of the domain there is a term which is interpreted onto it. Additionally, it can be assumed that any two constant symbols always denote different elements of the domain, this is termed *unique name assumption* (UNA). Clearly, these assumptions restrict models to domains which are isomorphic to the Herbrand domain, in a fashion similar to logic programming and deductive database semantics. These assumptions are almost normal in commonsense reasoning and we believe that there is no loss of generality w.r.t. to our purpose.

Lifschitz [4] has used the axiomatization [5] of explicit domains as an example of applying minimization to a subset of the theory, i.e. in the way NATs work. What is important is that explicit domains allow us to express orderings on the elements of the domain itself, a device needed to express value minimization. Consequently, in the rest of the paper we will tacitly assume that all theories include axioms for explicit domains and unique names.

**Definition 1** *(Explicit domain)*

*A NAT is said to have an explicit domain iff it contains axioms and blocks for DCA and UNA.*

Now we can proceed to discuss value minimization.

## 3 Definition of value minimization

In value minimization we compare values of functions relative to an ordering $\mathcal{R}$. This comparison is analogous to the comparison between extents of predicates relative to set inclusion in predicate circumscription. Let us introduce some notation: Let $T_{\mathcal{R}}(\Phi, z)$ be a theory where a function symbol $\Phi$ and function/predicate symbols in tuple $z$ appear as free variables. $T_{\mathcal{R}}(\Phi, z)$ also contains a definition of ordering $\mathcal{R}$. This ordering is assumed to be defined s.t. it is interpreted the same way in all models of the theory.

We will use the following notation with functions: let $\phi$, $\phi'$ be two function symbols, then

[5] See [*ibid*] for the axioms that have to be included in theories.

$$\phi \leq_{\mathcal{R}} \phi' \text{ stands for } \forall x(\mathcal{R}(\phi(x), \phi'(x)))$$

$$\phi <_{\mathcal{R}} \phi' \text{ stands for } (\phi \leq_{\mathcal{R}} \phi') \ \wedge \neg(\phi' \leq_{\mathcal{R}} \phi)$$

We use $\leq_{\mathcal{R},t_1,\ldots,t_n}$ to denote the same relation but with respect to a finite number $n$ of ground terms:

$$\phi \leq_{\mathcal{R},t_1,\ldots,t_n} \phi' \text{ stands for } \bigwedge_{i=1,\ldots,n}(\mathcal{R}(\phi(t_i), \phi'(t_i)))$$

Value minimization of a function $\phi$ with $Z$ varied is defined as follows:

$$T_{\mathcal{R}}(\phi, Z) \ \wedge \neg\exists\phi', z.[T_{\mathcal{R}}(\phi', z) \ \wedge \phi' <_{\mathcal{R}} \phi] \tag{1}$$

Sometimes it is necessary to minimize the value of a function relative to a finite number of ground terms. This was the case in the example where the term $Sit\_map(S_N)$ was minimized. In these cases we speak of *term minimization*. We define minimization of $\phi$ relative to terms $t_1, \ldots, t_n$ as follows:

$$T_{\mathcal{R}}(\phi, Z) \ \wedge \neg\exists\phi', z(T_{\mathcal{R}}(\phi', z) \ \wedge \phi' <_{\mathcal{R},t_1,\ldots,t_n} \phi)$$

Furthermore, since we only consider explicit domains, there is an even simpler formula for term minimization when the list $Z$ is empty. Let $\overline{c}$ stand for a tuple of ground terms $c_1, \ldots, c_k$, not containing the constant $\phi$, where $k$ is the number of terms relative to which we want to minimize $\phi$.

Let $T_{\mathcal{R}}(\overline{c})$ stand for $T_{\mathcal{R}} \bigwedge_{i=1,\ldots,k} \phi(t_i) = c_i$. Then we can write term minimization as:

$$T_{\mathcal{R}}(\overline{c}) \ \wedge \left[\forall\overline{c}' \left(T_{\mathcal{R}}(\overline{c}') \bigwedge_{i=1,\ldots,k} \mathcal{R}(c'_i, c_i)\right) \supset \left(\bigwedge_{i=1,\ldots,k} \mathcal{R}(c_i, c'_i)\right)\right]$$

## 4 Model-theoretic definition

In this section we define value minimization in model-theoretic terms[6]. Let us first introduce the following notation. Let $\mathcal{I}[[\pi]]$ stand for the set of tuples which belong to the extent of predicate $\pi$ in interpretation $\mathcal{I}$. For functions we use $\mathcal{I}[[\phi]]$ to denote the codomain of $\phi$ in $\mathcal{I}$. Also, we use $\mathcal{I}[[\phi]](\tau)$ to denote the object which —according to interpretation $\mathcal{I}$— function $\phi$ maps term $\tau$

[6] Please refer to Lifschitz [4] and Shoham [4] for a general discussion and results on preferential models.

into. Since we are considering Herbrand models only, we can have expressions like $\mathcal{I}[[\phi]](\tau) = \nu$ where $\nu$ is a term of the language of the object theory.

**Definition 2** *(Value-minimal) Let $T$ be a theory, let $\phi$ be a function and $Z$ a tuple of predicate/function constants in the language of $T$. Let $\mathcal{R}$ be a partial order defined over the elements of the universe. For two structures $\mathcal{M}$ and $\mathcal{M}'$ of $T$, we say $\mathcal{M} \leq^{(\phi,\mathcal{R});Z} \mathcal{M}'$ if*

*(i)* $|\mathcal{M}\ | = |\mathcal{M}'|$;
*(ii)* $\mathcal{M}[[\sigma]] = \mathcal{M}'[[\sigma]]$ *For each constant* $\sigma$ *s.t.* $\sigma \neq \phi$, $\sigma \notin Z$;
*(iii)* $\forall x.\mathcal{M}[[\phi]](x)\ \mathcal{R}\ \mathcal{M}'[[\phi]](x)$.

*A model $\mathcal{M}$ of $T$ is minimal relative to $\leq^{(\phi,\mathcal{R});Z}$ if there is no model $\mathcal{M}'$ of $T$ such that $\mathcal{M}' <^{(\phi,\mathcal{R});Z} \mathcal{M}$.*

In the case of *term minimization*, condition *(iii)* in Definition 2 becomes

$$\bigwedge_{i=1,\dots,k} \mathcal{M}[[\phi]](t_i)\ \mathcal{R}\ \mathcal{M}'[[\phi]](t_i)$$

and we write minimality as relative to the ordering $\leq^{(\phi,t_1,\dots,t_n,\mathcal{R});Z}$ between structures.

**Example 2** *(Exponentiation) Consider the following two interpretations of function "power" on the domain of naturals with $\leq$ as the usual "less than or equal" relation. By abuse of language, we define:*

$$\mathcal{M}[[power]](x) = x^2,\ \ \mathcal{M}'[[power]](x) = x^3.$$

*It is easy to establish that for all $x$ $\mathcal{M}[[power]](x) \leq \mathcal{M}'[[power]](x)$ and $\mathcal{M} <^{(power,\leq)} \mathcal{M}'$*

Finally, let us proceed to state that models of formula (1) are all and only those which are minimal w.r.t. $\leq^{(\phi,\mathcal{R});Z}$. This result is the counterpart of fundamental Proposition 2.5.1 of [4].

**Proposition 1** *A structure $\mathcal{M}$ is a model of*

$$T_{\mathcal{R}}(\phi, Z)\ \wedge \neg\exists\phi', z.[T_{\mathcal{R}}(\phi', z)\ \wedge \phi' <_{\mathcal{R}} \phi]$$

*if and only if $\mathcal{M}$ is a model of $T_{\mathcal{R}}(\phi, Z)$ and it is minimal relative to $\leq^{(\phi,\mathcal{R});Z}$.*
$\square$

## 5 Value minimization using NATs

To see why value minimization in NATs is simpler and more structured than the formulation in Section 3, let us start by observing how Definition 2 above concerns minimization relative to a fixed, external criterion $\mathcal{R}$, which is a partial ordering between elements of the universe.

The key feature of NATs is that an axiomatization of $\mathcal{R}$, i.e. basically the *closure* under reflexivity, transitivity and anti-symmetry, can be added as a block without much complications. This involves minimizing $\mathcal{R}$ w.r.t. its axiomatization [7].

A formalization of $\mathcal{R}$ and value-minimization in a form of circumscription other than NATs is bound to be more cumbersome since undesired side effects are likely to occur, and thus careful prioritization would be needed.

**Definition 3** *(Term ordering)*

*We say that a NAT $T$* has a term ordering $\mathcal{R}$ *with respect to a function $\phi$ if it contains a block defining a partial ordering $\mathcal{R}$ on the elements of the codomain of $\phi$, i.e. For all models $\mathcal{M}$ of $T$, and for all $x$ and $y$,*

$$\mathcal{M}[[\phi]](x)\ \mathcal{R}\ \mathcal{M}[[\phi]](y) \text{ iff } (\mathcal{M}[[\phi]](x), \mathcal{M}[[\phi]](y)) \in \mathcal{M}[[\mathcal{R}]].$$

Now, let us show that for each value minimization criterion $\leq^{(\phi,\mathcal{R});Z}$ an equivalent NAT formulation is found.

**Theorem 1** *(Value-minimal equivalence)*

*Let $T$ be an explicit domain NAT including term ordering $\mathcal{R}$ and let $\phi$ be a function constant and $Z$ be a tuple of predicate/function constants in the language of $T$. Then, a model $\mathcal{M}$ of $T$ is a model of*

$$\begin{array}{l} \{\phi, Z\ : \\ \qquad \forall x, y.\ \mathcal{R}(y, \phi(x)) \supset Ab(x, y) \\ \qquad T \\ \} \end{array} \tag{2}$$

*if and only if $\mathcal{M}$ is minimal relative to $\leq^{(\phi,\mathcal{R});Z}$.*

[7] Notice that there is no extra complexity involved in this process, since transitive closures are always defined as a second-order theory, as indeed the block defining $\mathcal{R}$ is.

**Proof:**

Consider a theory $T$ as described in the above proposition. Let $T_{\phi,\mathcal{R}}$ be the NAT of the form (2) built from $T$.

$(\Leftarrow)$

Assume $\mathcal{M}$ to be a model of $T$ minimal relative to $\leq^{(\phi,\mathcal{R});Z}$. Let us define the structure $\mathcal{M}_{Ab}$ by *augmenting* $\mathcal{M}$ as follows:

$$\mathcal{M}_{Ab}[[Ab]] = \{(x,y) : (y, \phi(x)) \in \mathcal{M}[[\mathcal{R}]]\}$$

Since $\mathcal{M}_{Ab}$ is a model of $T$, $\mathcal{M}_{Ab}$ is a model of $T_{\phi,\mathcal{R}}$ iff the extent of $Ab$ is minimal. We now show $\mathcal{M}_{Ab}$ is a model of $T_{\phi,\mathcal{R}}$ by contradiction.

Suppose there is a structure $\mathcal{M}'$ of $T_{\phi,\mathcal{R},Z}$ such that $|\mathcal{M}'| = |\mathcal{M}_{Ab}|$, $\mathcal{M}'[[\sigma]] = \mathcal{M}_{Ab}[[\sigma]]$ for every constant $\sigma$ different from $Ab$ and $\phi$ and not belonging to $Z$, and $\mathcal{M}'[[Ab]] \subset \mathcal{M}_{Ab}[[Ab]]$. We are going to show that $\mathcal{M}'$ cannot be a model of $T_{\phi,\mathcal{R},Z}$, i.e. $\mathcal{M}_{Ab}$ is a minimal model of $T_{\phi,\mathcal{R},Z}$. Since for all $x$

$$\mathcal{M}'[[\phi]](x)\ \mathcal{R}\ \mathcal{M}'[[\phi]](x) \Rightarrow (x, \mathcal{M}'[[\phi]](x)) \in \mathcal{M}'[[Ab]]$$

from above and the definition of $\mathcal{M}_{Ab}$ we can conclude that $(x, \mathcal{M}'[[\phi]](x)) \in \mathcal{M}_{Ab}[[Ab]]$. Now, since it follows that $\forall x.\mathcal{M}'[[\phi]](x)\ \mathcal{R}\ \mathcal{M}[[\phi]](x)$ we have established that $\mathcal{M}' \leq^{(\phi,\mathcal{R});Z} \mathcal{M}$. Also, given that $\mathcal{M}'[[Ab]] \subset \mathcal{M}_{Ab}[[Ab]]$, there exist $x, y$ s.t.

$$(x,y) \notin \mathcal{M}'[[Ab]] \text{ and } (x,y) \in \mathcal{M}_{Ab}[[Ab]].$$

For the same $x$ and $y$, since by transitivity:

$$\mathcal{M}[[\phi]](x)\ \mathcal{R}\ \mathcal{M}'[[\phi]](x) \Rightarrow y\ \mathcal{R}\ \mathcal{M}'[[\phi]](x)$$

and $y\ \mathcal{R}\ \mathcal{M}'[[\phi]](x)$ implies $(x,y) \in \mathcal{M}'[[Ab]]$, it follows that $\mathcal{M} \not\leq^{(\phi,\mathcal{R});Z} \mathcal{M}'$. This implies $\mathcal{M}$ is not minimal relative to $\leq^{(\phi,\mathcal{R});Z}$ which is contrary to the hypothesis. Therefore, $\mathcal{M}_{Ab}$ is a model of $T_{\phi,\mathcal{R}}$.

$(\Rightarrow)$

Suppose $\mathcal{M}$ is a model of $T_{\phi,\mathcal{R},Z}$. By definition of NATs, $\mathcal{M}$ is also a model of $T$. We show $\mathcal{M}$ is a model of $T$ minimal relative to $\leq^{(\phi,\mathcal{R});Z}$ by contradiction.

To do so, suppose there is a model $\mathcal{M}'$ of $T$ s.t.

$$\mathcal{M}' <^{(\phi,\mathcal{R});Z} \mathcal{M} \tag{3}$$

As before, we define $\mathcal{M}'$ by augmenting $\mathcal{M}$ by a predicate $Ab$ defined as follows:

$$\mathcal{M}'[[Ab]] = \{(x, y) : (y, \phi(x)) \in \mathcal{M}'[[\mathcal{R}]]\}$$

¿From (3), for all $x$, $\mathcal{M}'[[\phi]](x)\ \mathcal{R}\ \mathcal{M}[[\phi]](x)$ and –by transitivity– for all $x, y$:

$$y\ \mathcal{R}\ \mathcal{M}'[[\phi]](x)\ \Rightarrow\ y\ \mathcal{R}\ \mathcal{M}[[\phi]](x)$$

Thus, $(x, y) \in \mathcal{M}'[[Ab]]$ implies $(x, y) \in \mathcal{M}[[Ab]]$ and therefore

$$\mathcal{M}'[[Ab]] \subseteq \mathcal{M}[[Ab]].$$

On the other hand, from the strict order of (3) it follows that there is at least one $x$ for which the condition $\mathcal{M}[[\phi]](x)\ \mathcal{R}\ \mathcal{M}'[[\phi]](x)$ does not hold. For such an $x$, $(x, \mathcal{M}[[\phi]](x)) \notin \mathcal{M}'[[Ab]]$. However, for all $x$, $\mathcal{M}[[\phi]](x)\ \mathcal{R}\ \mathcal{M}[[\phi]](x)$ holds by definition of $\mathcal{R}$. Hence, $(x, \mathcal{M}[[\phi]](x)) \in \mathcal{M}[[Ab]]$ and we obtain

$$\mathcal{M}[[Ab]] \not\subseteq \mathcal{M}'[[Ab]]$$

Consequently, $\mathcal{M}$ is not a model of $T_{\phi,\mathcal{R},Z}$, which contradicts our assumption. As a result, $\mathcal{M}$ is minimal relative to $\leq^{(\phi,\mathcal{R});Z}$. □

### *5.1 Term Minimality*

Let us now move on to define minimization of ground terms. As before, consider an explicit domain NAT $T$ with term ordering $\mathcal{R}$, and let $\phi$ be a function constant, $t_1, \ldots, t_n$ be ground terms, and $Z$ be a tuple of predicate/function constants in the language of $T$. The following result is easily established.

**Proposition 2** *(Term-minimal equivalence)*

*A model $\mathcal{M}$ of $T$ is a model*[8] *of*

$$\begin{array}{l}\{\phi, Z \ : \\ \qquad \forall y.\ \mathcal{R}(y,\phi(t_1)) \supset Ab(t_1,y) \\ \qquad \ldots \\ \qquad \forall y.\ \mathcal{R}(y,\phi(t_n)) \supset Ab(t_n,y) \\ \qquad T \\ \}\end{array} \tag{4}$$

*if and only if $\mathcal{M}$ is minimal relative to $\leq^{(\phi,t_1,\ldots,t_n,\mathcal{R});Z}$.* □

## 6 Final Remarks

(i) The straightforward results obtained here depend crucially on the possibility to define a partial order on terms independently from the interpretation. This requires axiomatizing a) explicit domains and b) partial orders, both of which require predicate circumscription to yield the expected results. The NAT syntax allows us to nest these as subblocks of the theory of interest.

(ii) The blocks implementing value-minimization of a function are just another kind of block, and therefore can be part of another NAT. It was only for the sake of definition that we described them as outermost w.r.t. theories.

(iii) The ordering $\leq^{(\phi,\mathcal{R});Z}$ between structures in Definition 2 can be extended to value minimization of a set of functions and to incorporate possible priorities between them by following the standard approach [4].

(iv) Finally, let us point out how Lifschitz's NAT notation is readily adapted to our definitions. We will write

$$\{C_1,\ldots,C_m,\ min_{\mathcal{R}}\ \phi : A_1,\ldots,A_n,A_{\mathcal{R}}\} \tag{5}$$

where $A_{\mathcal{R}}$ defines[9] $\mathcal{R}$— to denote blocks of the form $\{C_1,\ldots,C_m,\phi : \mathcal{R}(y,\phi(x)) \supset Ab(x,y), A_1,\ldots,A_n,A_{\mathcal{R}}\}$

Intuitively, block (5) refers to a theory consisting of blocks $A_1,\ldots,A_n$ and the block $A_{\mathcal{R}}$ defining $\mathcal{R}$, and where value minimization of function $\phi$ is performed while predicate/function constants $C_1,\ldots,C_m$ are varying.

[8] For $n=1$ we just have $Ab(y)$ in the right hand side of the axiom.

[9] We are assuming that $A_{\mathcal{R}}$ is defined using an NAT block such that statements about $\mathcal{R}$ outside of $A_{\mathcal{R}}$ do not affect the definition of $\mathcal{R}$.

## References


[1] C. Baral, A.Gabaldon and A. Provetti. Formalizing Narratives using Nested Circumscription. In Proc. of *14th AAAI National Conference on Artificial Intelligence* (1996) 652–657.

Also in the Working notes of the 3rd symposium *Logical Formalizations of Commonsense Reasoning,* 15–24.

[2] C. Baral, M. Gelfond and A. Provetti. Representing Actions: Laws, Observations and Hypothesis.

To appear on *Journal of Logic Programming,* special issue on actions (1996). Also available via *http://cs.utep.edu/chitta/chitta.html*

An earlier version appeared in the Working Notes of AAAI Spring Symposium *Extending Theories of Actions: Formal Theories and Practical Applications.*

[3] A. Baker. Nonmonotonic reasoning in the framework of situation calculus. *Artificial Intelligence*, **49** (1991), 5–23.

[4] B. Grosof. Generalizing Priorization. In Proc. of *KR - Principles of Knowledge Representation and Reasoning* (Morgan Kauffman, 1991) 289–300.

[5] N. Kartha and V. Lifschitz. Actions with indirect effects (preliminary report). In Proc. of *KR - Principles of Knowledge Representation and Reasoning*, (Morgan Kauffman, 1994) 341–350.

[6] V. Lifschitz. Circumscription. In *Handbook of Logic in Artificial Intelligence and Logic Programming,* (Oxford Un. Press, 1993) 297–352.

[7] V. Lifschitz. Nested Abnormality Theories. *Artificial Intelligence*, **74** (1995), 351–365.

[8] F. Lin and R. Reiter. How to Progress a Database II: the STRIPS Connection. *Proc. the of 14th IJCAI Conference*, (1995) 2001–2007.

[9] J. McCarthy. Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence,* **26**(3), (1986) 89–116.

[10] R. Reiter. Equality and domain closure in in first-order data bases. In *J. of the ACM* **27**(2), (1980) 235–249.

[11] Y. Shoham. Nonmonotonic logics: meaning and utility. *Proc. the of 10th IJCAI Conference*, (1987) 388–393.

## A Overview of Nested Circumscription

This Appendix is a resume of Lifschit'z paper in AI Journal and has been included for referee' convenience.

Circumscription [4,4] with Nested Abnormality Theories (NATs) is a novel circumscription technique introduced in [4]. NATs allow the use of several abnormality predicates to specify a body of common-sense knowledge without their circumscription conflicting. Lifschitz introduces the notion of *block:* a set of axioms $A_1, \ldots, A_n$—possibly containing the abnormality predicate $Ab$—"describes" a set of predicates/constants $C_1, \ldots, C_m$, which are allowed to vary during circumscription of $Ab$. The notation for such a theory is

$$C_1, \ldots, C_m \ : \ A_1, \ldots, A_n. \tag{A.1}$$

where each $A_i$ may itself be a block of form (A.1). A block may be viewed as a set of axioms describing a collection of predicates/constants, possibly depending on other descriptions in embedded blocks. Block (A.1) can be expressed in terms of the circumscription operator as follows:

$$\mathrm{CIRC}[A_1 \ \wedge \ldots \ \wedge A_n; Ab; C_1, \ldots, C_m]. \tag{A.2}$$

In addition to NATs, Lifschitz [4] introduces the idea of replacing the predicate *"Ab"* by an existentially quantified variable. Noticing that $Ab$ plays an auxiliary role and that the interesting consequences of the theory are those which *do not* contain $Ab$, $Ab$ is replaced by an existentially quantified variable. To put it differently, if $F(Ab)$ denotes (A.2), and *"ab"* is a predicate variable of the same arity as $Ab$, we are interested in the consequences of the sentence $\exists ab F(ab)$. The effect of this modification is that abnormality predicates become local to the block where each of them is used.

### *A.1 Syntax and semantics of NATs*

The following definitions are from [4]. Let $L$ be a second order language which does not include $Ab$. For every natural number $k$, let $L_k$ be the language obtained by adding the $k$-ary predicate constant $Ab$ to $L$. $\{C_1, ..., C_m : A_1, ..., A_n\}$ is a *block* if each $C_1, \ldots, C_m$ is a predicate or a function constant of $L$, and each $A_1, \ldots, A_n$ is a formula of $L_k$ or a block.

A *Nested Abnormality Theory* is a set of blocks. The semantics of NATs is characterized by a mapping $\varphi$ from blocks into sentences of $L$. If $A$ is a formula of language $L_k$, $\varphi A$ stands for the universal closure of $A$, otherwise

$$\varphi\{C_1, ..., C_m \ : \ A_1, ..., A_n\} \ = \ \exists ab F(ab)$$

where

$$F(Ab) = \mathrm{CIRC}[\varphi A_1 \wedge ... \wedge \varphi A_n; Ab; C_1, ..., C_m]$$

For any NAT $T$, $\varphi T$ stands for $\{\varphi A | A \in T\}$. A *model* of $T$ is a model of $\varphi T$ in the sense of classical logic. A *consequence* of $T$ is a sentence $\phi$ of language $L$ that is true in all models of $T$. In this paper, as suggested in [4], we will use the abbreviation

$$\{C_1, \ldots, C_m, \ min \ P : A_1, \ldots, A_n\}$$

to denote blocks of the form

$$\{C_1, \ldots, C_m, P : P(x) \supset Ab(x), A_1, \ldots, A_n\}$$

As the notation suggests, this type of block is used when it is necessary to circumscribe a particular predicate $P$ in a block. In [4] it is shown that

$$\varphi\{C_1, \ldots, C_m, \ min \ P : A_1, \ldots, A_n\}$$

is equivalent to the formula

$$\mathrm{CIRC}[A_1 \wedge \ldots \wedge A_n; P; C_1, \ldots, C_m]$$

when each of $A_i$ is a sentence.

### *A.2 Axiomatization of orderings*

As an example, consider the NAT definition of subsequence [10] we had in [4] to order terms like

[10] Recall that $\alpha = A_1, \ldots, A_m$ is a subsequence of $\beta = B_1, \ldots, B_n$ if there exists a strictly increasing sequence $i_1, \ldots, i_m$ of indices of $\beta$ such that for all $j = 1, \ldots, m$, we have $A_j = B_{i_j}$.

$$A_n \ldots \circ A_2 \circ A_1 \circ \epsilon$$

read as "$A_1$ then $A_2$ then … $A_n$".

**Example 3** *(Defining subsequences by a NAT)*

$$
\begin{aligned}
&B_{Subsequence} = \\
&\{min\ Subsequence\ : \\
&\qquad Subsequence(\epsilon, \epsilon) \\
&\qquad Subsequence(\alpha, \alpha_1) \supset Subsequence(\alpha, a \circ \alpha_1) \\
&\qquad Subsequence(\alpha, \alpha_1) \supset Subsequence(a \circ \alpha, a \circ \alpha_1) \\
&\}
\end{aligned}
$$

Of course, $B_{Subsequence}$ is a second order theory, irrespective from the NAT definition. For the sake of our investigation, we are specially interested in NATs for which an equivalent first-order formalization can be given. The idea is to *simplify* a theory of nested blocks by substituting the innermost blocks with their equivalent first-order theories, and repeat the process inside out. The final first-order theory —if it exists— can be used for deduction or comparison purposes.