# Narcissists, Stepmothers and Spies

Maarten Marx*
LIT, ILLC, Universiteit van Amsterdam, The Netherlands
Email: marx@science.uva.nl, www.science.uva.nl/∼marx

**Abstract**

This paper investigates the possibility of adding machinery to description logic which allows one to define self-referential concepts. An example of such a concept is a narcissist, someone who loves himself. With domains in which the natural ontology is a graph instead of a tree, this extra expressive power is often desired (e.g., when writing an ontology about web pages or molecular structures). Our results show that one has to be very careful with such additions. We add self-reference to $\mathcal{ALC}$ with inverse. Then we obtain all well known difficulties of having individual concepts or nominals together with inverse relations and even worse, checking for concept consistency becomes undecidable. Most of this expressive power seems not to be needed and we can identify a useful fragment whose complexity does not exceed that of $\mathcal{ALC}$.

## 1 Introduction and Motivation

We investigate adding a form of self-reference to description logic. This form is inspired by the downarrow operator from hybrid logic which names the "here and now" [3]. Let us first look at the natural language definitions of the first two concepts in the title:

**narcissist:** *someone who loves* oneself;

**stepmother:** *a female who is married to a person who has a child which is not* hers.

An example from the web could be a "solipsistic page" —a page which only links to pages which link back to *it*.

Using the notions of bisimulations for description logics developed by Kurtonina and de Rijke [7] one can simply show that these concepts are not definable

*Research supported by NWO grants 612-062-001 and 612.000.106. I thank the anonymous referee for useful comments.

in $\mathcal{ALC}$ and other extensions. This is due to the fact that these concepts exploit the graph like structure of the underlying domain, while $\mathcal{ALC}$ concepts can only capture (part of) the tree like aspects of it. This part of the design makes it robustly decidable [9].

There are some domains in which the graph like nature of the relations is important and the definition of concepts makes use of it. The web is a good example. One has to be careful in designing languages which may speak about the graph like nature. Once grids can be defined, undecidability is very close.

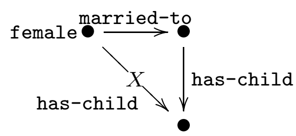
Instead of adding variables as in hybrid logic, we add here the personal pronouns **I** and **me** to description logic with the following intended meaning:

> If $C$ is a (complex) concept and $a$ an element of the domain, then $a$ belongs to $\mathbf{I}.C$ if $a$ belongs to $C$ under the assumption that all occurrences of **me** in $C$ denote the individual concept $\{a\}$.

Note that **me** can be seen as a kind of dynamic version of the one-of operator. With **I** and **me** we can define the earlier mentioned concepts:

| | |
|---|---|
| **narcissist** | $\mathbf{I}.\exists\ \texttt{loves}\ \mathbf{me}$ |
| **stepmother** | $\texttt{female} \sqcap \mathbf{I}.\exists\ \texttt{married-to}\ \exists\ \texttt{has-child}\ \neg\exists\ \texttt{has-child}^{-1}\ \mathbf{me}$ |
| **solipsitic web page** | $\mathbf{I}.\forall\ \texttt{has-link}\ \exists\ \texttt{has-link}\ \mathbf{me}$. |

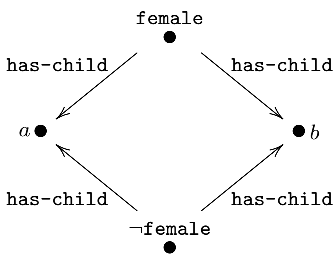
The definition of stepmother can graphically be represented as below. Here the node labeled `female` is the stepmother.




We can also define concepts which intrinsically need three variables, like sibling. (This definition goes back to Schröder.)

**sibling:** $\mathbf{I}.\exists\ \texttt{has-child}^{-1}\ (\texttt{female} \sqcap \exists\ \texttt{has-child}\ (\neg\,\mathbf{me} \sqcap \exists\ \texttt{has-child}^{-1}\ (\neg\,\texttt{female} \sqcap \exists\ \texttt{has-child}\ \mathbf{me})))$.

In the picture below, the (different) nodes $a$ and $b$ are both siblings.



From DL and hybrid logic it is known that the combination of nominals or individual concepts together with inverse roles makes reasoning much harder (e.g., checking concept consistency for empty T-Box in $\mathcal{ALCI}$ with one nominal is EXPTIME hard). With self-reference things are much worse. Consider a T-Box $\{\top \sqsubseteq (1)\}$, with (1) the concept:

$$\begin{array}{rr} & \exists\,\texttt{has-link}\,\top \quad \sqcap \\ (1) & \forall\,\texttt{has-link}\,\neg\texttt{startpage} \quad \sqcap \\ & \mathbf{I}.\forall\,\texttt{has-link}\,\forall\,\texttt{has-link}^{-1}\,\mathbf{me}. \end{array}$$

The concept $\texttt{startpage}$ can only be non empty in a model making this T-Box true if the model is infinite. This is because $\mathbf{I}.\forall\,\texttt{has-link}\,\forall\,\texttt{has-link}^{-1}\,\mathbf{me}$ expresses that *has-link* is an injective relation.

Things get even worse with "spy pages": a web page that has a direct link to all pages which can be reached by following a path of links from itself. They are defined by

$$(2) \qquad \mathbf{I}.\forall\,\texttt{has-link}\,\forall\,\texttt{has-link}\,\exists\,\texttt{has-link}^{-1}\,\mathbf{me}.$$

Now consider the concept (3), in which $R$ is a new relation.

$$\begin{array}{rr} & \mathbf{I}.\forall\,\texttt{R}\,\forall\,\texttt{has-link}\,\exists\,\texttt{R}^{-1}\,\mathbf{me} \quad \sqcap \\ (3) & \exists\,\texttt{R}\,\texttt{startpage} \quad \sqcap \\ & \forall\,\texttt{R}\,(1). \end{array}$$

The previous result immediately implies that the concept defined by (3) is only non empty on infinite models.

In the remainder of the paper we show undecidability of concept consistency with empty T-Box for $\mathcal{ALC}$ with $\mathbf{I}$ and $\mathbf{me}$ added. For this result only one relation is needed and no converse. This contradicts the decidability result from Theorem 7.10 in [2]. The mistake in that proof lies in the given reduction to a problem without occurrences of $\mathbf{I}$. The concept (1) above showed that this cannot be done in a finite way.

On the positive side, we also define a decidable existential version of the language. Here we tame the power of the $\mathbf{I}$–$\mathbf{me}$ construction by disallowing universal quantifiers in the scope of $\mathbf{I}$ (as occurring in (1) and (2)).

Before we start let's make things precise. If $\mathcal{X}$ denotes some description language, let $\mathcal{X}\mathsf{self}$ be the language with the added clauses:

- $\mathbf{me}$ is a concept;
- if $C$ is a concept, then $\mathbf{I}.C$ is a concept.

Here $C$ is called the *scope* of $\mathbf{I}$. The concept **me** occurs *free* in $C$ if it is not in the scope of some $\mathbf{I}$. We only consider concepts in which every occurrence of **me** occurs in the scope of some $\mathbf{I}$. Now we can make the meaning of the new concepts precise.

> $a$ belongs to $\mathbf{I}.C$ if $a$ belongs to $C$ under the assumption that all free occurrences of **me** in $C$ denote the individual concept $\{a\}$.

In the sequel we use $C \to D$ as an abbreviation of $\neg C \sqcup D$.

## 2 The existential fragment

In this section we tame the power of self reference by restricting the type of concepts which can occur in the scope of $\mathbf{I}$. Decidability and a matching complexity bound are obtained by a translation into the guarded fragment of first order logic with three variables [1, 6].

Note that the spy-page (2) is equivalent to the first order formula $\forall y(Rxy \to \forall z(Ryz \to Rxz))$, when $R$ is interpreted as *has-link*. This formula uses three variables in a nonreducible way and is not equivalent to a (loosely) guarded formula.

The description logic we consider is $\mathcal{ALCI}$, $\mathcal{ALC}$ with inverse roles. Every $\mathcal{ALCI}$ concept is equivalent to one in negation normal form, that is, constructed by $\sqcap, \sqcup, \exists\, \mathtt{R}, \forall\, \mathtt{R}, \exists\, \mathtt{R}^{-1}$ and $\forall\, \mathtt{R}^{-1}$ from atomic concepts and their negations. We define the *existential* $\mathcal{ALCI}\mathsf{self}$ concepts as those constructed by $\sqcap, \sqcup, \exists\, \mathtt{R}, \exists\, \mathtt{R}^{-1}$ and $\mathbf{I}$ from atomic concepts, **me**, their negations and $\neg\exists\, \mathtt{R}\ \mathbf{me}$ and $\neg\exists\, \mathtt{R}^{-1}\ \mathbf{me}$. Thus universal quantification is not allowed, except in the form of $\neg\exists\, \mathtt{R}\ \mathbf{me}$ in which form it is just an atomic statement.

The set of $\mathcal{ALCI}\mathsf{self}^{\exists}$ concepts is the smallest set such that every atomic concept name including **me** and their negations are concepts, and if $C$ and $D$ are concepts, then $C \sqcap D, C \sqcup D, \exists\, \mathtt{R}\, C, \exists\, \mathtt{R}^{-1}\, C, \forall\, \mathtt{R}\, C, \forall\, \mathtt{R}^{-1}\, C$ are also concepts. Moreover, if $C$ is an existential $\mathcal{ALCI}\mathsf{self}$ concept, then $\mathbf{I}.C$ is also $\mathcal{ALCI}\mathsf{self}^{\exists}$ concept. An $\mathcal{ALCI}\mathsf{self}^{\exists}$ T–Box consists of a set of GCI's of the form $C \sqsubseteq D$, for $C, D$ $\mathcal{ALCI}\mathsf{self}^{\exists}$ concepts with the requirement that $\mathbf{I}$ does not occur in $C$. Note that $\mathcal{ALCI}\mathsf{self}^{\exists}$ contains $\mathcal{ALCI}$ and that narcissist, stepmother and sibling can still be defined. Also note that even if $\mathcal{ALCI}\mathsf{self}^{\exists}$ is not closed under negation, the subsumption problem can in specific cases still be reduced to the satisfiability problem. In particular, if $\neg D$ is equivalent to an $\mathcal{ALCI}\mathsf{self}^{\exists}$ concept, then $\Sigma \models C \sqsubseteq D$ reduces to the satisfiability problem coverd by the following theorem.

**Theorem 1** *Let $\Sigma$ be an $\mathcal{ALCI}\mathsf{self}^{\exists}$ T–Box and $C$ an $\mathcal{ALCI}\mathsf{self}^{\exists}$ concept. The problem of checking concept consistency ($\Sigma \not\models C \doteq \bot$) is decidable in* EXPTIME.

Proof. We translate the problem using the standard translation to the universal guarded fragment[1] with three variables. In this fragment only universal formulas need to be guarded. Grädel [6] showed that the satisfiability problem for this fragment is complete for exponential time. The $\mathcal{ALCI}$ concepts are translated as usual (e.g., as in Table 2 of Borgida [5]). The new clauses are

$$\begin{array}{lcllcl} \mathcal{T}^x(\mathbf{I}.C) & := & \exists w(x = w \wedge \mathcal{T}^x(C)) & \mathcal{T}^y(\mathbf{I}.C) & := & \exists w(y = w \wedge \mathcal{T}^y(C)) \\ \mathcal{T}^x(\,\mathbf{me}) & := & x = w & \mathcal{T}^y(\,\mathbf{me}) & := & y = w. \end{array}$$

For example, $\mathcal{T}^x(\mathbf{I}.\exists\ \mathtt{R}\,\exists\ \mathtt{R}\ \mathbf{me})$ is $\exists w(w = x \wedge \exists y(Rxy \wedge \exists x(Ryx \wedge x = w)))$ which is equivalent to $\forall y(Rxy \rightarrow Ryx)$. The restriction on the scope of the $\mathbf{I}$ ensures that there are no non guarded universal quantifiers in the translation.

Because of the restriction on the form of the CGI's in the T–box, we may assume that they all have the form $C \sqsubseteq \top$, for $C$ and $\mathcal{ALCI}\mathsf{self}^{\exists}$ concept. Then $\Sigma \not\models C \doteq \bot$ iff the universally guarded sentence $\forall x(x = x \rightarrow \bigwedge\{\mathcal{T}^x(C) \mid C \sqsubseteq \top \in \Sigma\}) \wedge \exists x \mathcal{T}^x(C)$ is satisfiable. QED

# 3 Undecidability

**Theorem 2** *Let $C$ be an $\mathcal{ALC}\mathsf{self}$ concept containing just one relation symbol $R$. The problem of checking concept concistency with empty T-Box for such $C$ is undecidable.*

Undecidability is shown by encoding the $\mathbb{N} \times \mathbb{N}$ tiling problem (cf. [4]). The main point in such a proof is to show that two commuting functions *up* and *right* can be defined. Let $\Sigma$ be a T-Box consisting of the following CGI's:

$$\begin{array}{lcl} \top & \sqsubseteq & \exists\ \mathtt{up}\,\top \sqcap \exists\ \mathtt{right}\,, \\ \top & \sqsubseteq & \mathbf{I}.\forall\ \mathtt{up}^{-1}\,\forall\ \mathtt{up}\ \mathbf{me} \sqcap \mathbf{I}\forall\ \mathtt{right}^{-1}\,\forall\ \mathtt{right}\ \mathbf{me}, \\ \top & \sqsubseteq & \mathbf{I}.\forall\ \mathtt{up}^{-1}\,\forall\ \mathtt{right}^{-1}\,\forall\ \mathtt{up}\,\forall\ \mathtt{right}\ \mathbf{me}. \end{array} \tag{4}$$

If we apply the standard translation from the previous section to (4) (and simplify formulas) we obtain a theory which says that for all $x$,

$$\begin{array}{l} \exists y R_{\mathsf{up}}xy \wedge \exists y R_{\mathsf{right}}xy, \\ \forall yz(R_{\mathsf{up}}xy \wedge R_{\mathsf{up}}xz \rightarrow y = z) \wedge \forall yz(R_{\mathsf{right}}xy \wedge R_{\mathsf{right}}xz \rightarrow y = z), \\ \forall yz(R_{\mathsf{up}}yx \wedge R_{\mathsf{right}}zy \rightarrow \forall w(R_{\mathsf{up}}zw \rightarrow R_{\mathsf{right}}wx)). \end{array}$$

Thus $I \models \Sigma$ if and only if *I(up)* and *I(right)* are commuting total functions. Having this part it is standard to code up the tiling problem.

[1] Formulas in this fragment are constructed from atoms and their negations by conjunction, disjunction, unrestricted existential quantification and guarded universal quantification [8].

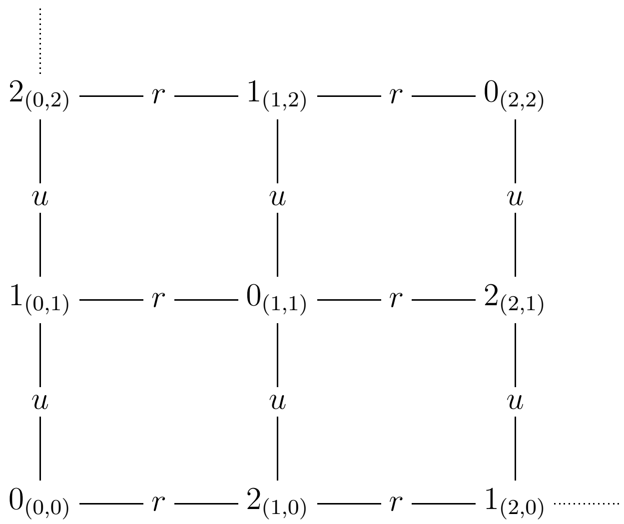


Figure 1: Our standard model of the grid.

Now we turn to the proof of the theorem. Because we do not want to use inverse relations and only one relation symbol, we need some additional coding. In Figure 1 we present how we would model the grid. In this model there is only one relation $R$ which is symmetric. The atomic concepts are $\{0, 1, 2\}$ and $\{u, r\}$. The nodes in the picture are given by the labels $u$ and $r$ and by the labels $i_{(k,m)}$. These last nodes correspond to positions in the grid. For $i \in \{0, 1, 2\}$, define $s(i) = i + 1 \bmod 3$ and $p(i) = i + 2 \bmod 3$. The idea is to model $\exists\, \mathtt{up}\, C$ by (for $i \in \{0, 1, 2\}$)

$$i \rightarrow \exists\, \mathtt{R}\, (u \sqcap \exists\, \mathtt{R}\, (s(i) \sqcap C)),$$

and $\exists\, \mathtt{right}\, C$ by $i \rightarrow \exists\, \mathtt{R}\, (r \sqcap \exists\, \mathtt{R}\, (p(i) \sqcap C))$. The corresponding relation $\mathtt{up}$ then is

$$\{\langle x, y\rangle \mid \exists z(x\mathtt{R}z\mathtt{R}y \wedge z \in I(u) \wedge \exists i(x \in I(i) \wedge y \in I(s(i))))\}.$$

We now present a number of concepts (5)–(9) which force an unravelled model to have the model from Figure 1 as a substructure in the case that each element in the domain belongs to these concepts. The first concept (5) expresses that the relation $R$ is symmetric.

(5) $$\mathbf{I}.\forall\, \mathtt{R}\, \exists\, \mathtt{R}\; \mathbf{me}.$$

For $i \in \{0, 1, 2\}$, (6) expresses that every point has an *up* and a *right* successor.

(6) $$i \rightarrow \exists\, \mathtt{R}\, (u \sqcap \exists\, \mathtt{R}\, s(i)) \sqcap \exists\, \mathtt{R}\, (r \sqcap \exists\, \mathtt{R}\, p(i)).$$

The next two concepts express that these successors are unique and that the relations $\mathtt{up}$ and $\mathtt{right}$ are irreflexive. For $i \in \{0, 1, 2\}$,

(7) $s(i) \rightarrow \mathbf{I}.\forall\, \mathtt{R}\, (u \rightarrow \forall\, \mathtt{R}\, (i \rightarrow (\neg\, \mathbf{me} \sqcap \forall\, \mathtt{R}\, (u \rightarrow \forall\, \mathtt{R}\, (s(i) \rightarrow \mathbf{me})))))$

(8) $p(i) \rightarrow \mathbf{I}.\forall\, \mathtt{R}\, (r \rightarrow \forall\, \mathtt{R}\, (i \rightarrow (\neg\, \mathbf{me} \sqcap \forall\, \mathtt{R}\, (r \rightarrow \forall\, \mathtt{R}\, (p(i) \rightarrow \mathbf{me}))))).$

$$\begin{tikzcd}
s(i) \arrow[r, dash, dashed, "r"] \arrow[d, dash, "u"'] & i \arrow[d, dash, dashed, "u"] \\
i \arrow[r, dash, "r"'] & p(i)
\end{tikzcd}$$

Figure 2: (9) expresses a confluence property.

The last concept expresses the confluence property depicted in Figure 2 (assuming $R$ is a symmetric relation). For $i \in \{0, 1, 2\}$,

$$\begin{aligned} s(i) \rightarrow \mathbf{I}.\forall\, \mathtt{R}\, (u \rightarrow \forall\, \mathtt{R}\, (i \rightarrow \forall\, \mathtt{R}\, (r \rightarrow \forall\, \mathtt{R}\, (p(i) \rightarrow \\ \exists\, \mathtt{R}\, (u \sqcap \exists\, \mathtt{R}\, (i \sqcap \exists\, \mathtt{R}\, (r \sqcap \exists\, \mathtt{R}\ \mathbf{me}))))))). \end{aligned} \tag{9}$$

Concepts (5)–(9) take care of the structural side of the encoding. Now let $T = \{T_1, \ldots, T_k\}$ be a set of tile types. For each type $T_i$, there is a corresponding atomic concept $t_i$. The next concept expresses that at every grid position exactly one tile concept holds. For $i \in \{0, 1, 2\}$,

$$i \rightarrow \bigsqcup_{1 \le n \le k} t_k \sqcap \bigwedge_{1 \le n \ne m \le k} \neg(t_n \sqcap t_m). \tag{10}$$

The last two concepts express that the colors of the tiles match. For $i \in \{0, 1, 2\}$ and $1 \le n \le k$,

$$i \sqcap t_n \rightarrow \forall\, \mathtt{R}\, (u \rightarrow \forall\, \mathtt{R}\, (s(i) \rightarrow \bigsqcup\{t_m \mid \mathit{top}(T_n) = \mathit{bottom}(T_m)\})) \tag{11}$$

$$i \sqcap t_n \rightarrow \forall\, \mathtt{R}\, (r \rightarrow \forall\, \mathtt{R}\, (p(i) \rightarrow \bigsqcup\{t_m \mid \mathit{right}(T_n) = \mathit{left}(T_m)\})). \tag{12}$$

Let $\phi_T$ be the conjunction of all concepts (5)–(12). It is straightforward to show that $T$ tiles the grid if and only if the T–box $\{\top \sqsubseteq \phi_T\}$ is satisfiable. So we showed that checking T-Box concistency is undecidable. To obtain the result in the theorem we use the spypoint technique from [3]. Consider the following concept

$$\begin{aligned} &\mathbf{I}.\forall\, \mathtt{R} \forall\, \mathtt{R} \exists\, \mathtt{R}\ \mathbf{me} \quad \sqcap \\ &\forall\, \mathtt{R} \forall\, \mathtt{R}\, \mathbf{I}.\forall\, \mathtt{R} \exists\, \mathtt{R}\ \mathbf{me}. \end{aligned} \tag{13}$$

Applying the standard translation, the meaning becomes clearer:

$$\forall yz(Rwy \wedge Ryz \rightarrow Rzw) \wedge \forall xyz(Rwy \wedge Ryz \wedge Rzx \rightarrow Rxz).$$

Together the conjuncts imply that $R$ is transitive from $w$: $\forall yz(Rwy \wedge Ryz \rightarrow Rwz)$. Thus we can use (13) to forces that every element in the model belongs to a concept.

Putting everything together we obtain that $T$ tiles the grid if and only if the concept (13) $\sqcap\ \exists\, \mathtt{R}\, 0\ \sqcap\ \forall\, \mathtt{R}\, \phi_T$ is non empty. This finishes the proof of Theorem 2.

## 4 Conclusion

We showed that adding a simple form of self reference to $\mathcal{ALC}$ makes it very, and indeed, too expressive: the language becomes undecidable. By restricting the concepts to which self reference can be applied we "tamed" the expressive power and obtained decidability. What is left is more or less the possibility of expressing that certain loops exist; still a useful extension of $\mathcal{ALC}$. I conjecture that existing tableau based procedures for $\mathcal{ALC}$ can be adapted to include this limited form of self reference. I even believe that the problem of checking for concept consistency with empty T-Box can still be done in PSPACE.

## References


[1] H. Andréka, J. van Benthem, and I. Németi. Modal languages and bounded fragments of predicate logic. *J. of Philosophical Logic*, 27(3):217–274, 1998.

[2] C. Areces. *Logic Engineering*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 2000.

[3] C. Areces, P. Blackburn, and M. Marx. Hybrid logics: Characterization, interpolation and complexity. *J. of Symbolic Logic*, 66(3):977–1010, 2001.

[4] E. Börger, E. Grädel, and Y. Gurevich. *The Classical Decision Problem*. Springer Verlag, 1997.

[5] A. Borgida. On the relative expressiveness of description logics and predicate logics. *Artificial Intelligence*, 82:353–367, 1996.

[6] E. Grädel. On the restraining power of guards. *Journal of Symbolic Logic*, 64(4):1719–1742, 1999.

[7] N. Kurtonina and M. de Rijke. Classifying description logics. In M.-C. Rousset et al, editor, *Proc. International Workshop on Description Logics (DL'97)*, pages 49–53, LRI, CNRS, Gif sur Yvette, 1997.

[8] M. Marx. Tolerance logic. *Journal of Logic, Language and Information*, 10:353–373, 2001.

[9] M. Vardi. Why is modal logic so robustly decidable? In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science 31*, pages 149–184. American Math. Society, 1997.