# A Formalization of Viewpoints [*]

Giuseppe Attardi [†]
Maria Simi [‡]

TR-93-062

October, 1993

## Abstract

We present a formalisation for the notion of *viewpoint*, a construct meant for expressing several varieties of relativised truth. The formalisation consists in a logic which extends first order predicate calculus through an axiomatization of provability and with the addition of proper reflection rules. The extension is not conservative, but consistency is granted. Viewpoints are defined as set of reified meta-level sentences. A proof theory for viewponts is developed which enables to carry out proofs of sentences involving several viewpoints. A semantic account of viewpoints is provided, dealing with issues of self referential theories and paradoxes, and exploiting the notion of *contextual entailment*. Notions such as beliefs, knowledge, truth and situations can be uniformly modeled as provability in specialised viewpoints, obtained by imposing suitable constraints on viewpoints.

---

[*]Part of this work has been done while the authors were visiting the International Computer Science Institute, Berkeley CA.

[†]Dipartimento di Informatica, Università di Pisa, Corso Italia 40, I-56125 Pisa, Italy. Net: attardi@di.unipi.it.

[‡]Dipartimento di Informatica, Università di Pisa, Corso Italia 40, I-56125 Pisa, Italy. Net: simi@di.unipi.it.

# 1   Introduction

Any adequate theory of knowledge representation must be capable of dealing with certain modal notions such as belief, knowledge and truth, or in general any notion of relativised truth. These notions are essential in particular for the reasoning of rational agents, for building micro-theories which can be exploited outside their original scope [Guha 91], for enabling programs to translate knowledge from one source into another.

Modal logic has been used to address propositional attitudes, while the formalisation of contexts [McCarthy 93] aims, among other things, at addressing the issues of relativised truth. In this paper we attempt to develop a theory of relativised truth which can cover also propositional attitudes.

First-order modal logics are not expressive enough for representing all natural statements involving knowledge and beliefs: sentences like *John knows that Bill knows something that he doesn't know*, or *John believes everything that Bill does* require quantification over propositions and properties. Approaches based on modal logic also suffer other drawbacks: issues related to the need of rigid designators, problems with opacity of quantification within the scope of a modal operator, the problem of logical omniscience.

An alternative approach to modal logic is to introduce a metalanguage, view propositions as objects and express modalities as first-order predicates. Such syntactic treatment of modalities has several appeals: great expressivity and extendibility is achieved without adding special operators to the language; well established first-order theorem proving techniques remain applicable; the complications of possible world semantics are avoided.

The advantages of a syntactic approach to the representation of truth, knowledge and beliefs have been largely discussed in the literature ([McCarthy 79], [Moore 77], [Konolige 82], [Perlis 85], [Perlis 88], [Davies 90]).

Along this line of research several proposals are based on extending a logic calculus with a notion of provability. Provability in an agent's theory is considered as one of the primitive relations for the formalization of beliefs and knowledge.

These proposals differ in the degree of connection between object-theory and meta-theory, ranging from a semantic connection of completely separate theories as in [Konolige 82], to the reflection principles of FOL [Weyhrauch 80], bridging two still distinct theories, to the proposal of a single amalgamated theory encompassing object and meta-level, but with weak reflection rules [Bowen 82].

Nevertheless, a satisfactory first order theory of relativised truth is not easy to develop since one must face delicate issues of semantics and must avoid the pitfalls of paradoxes arising from self referential sentences, which trickle in by diagonalization [Montague 63]. One may think that these sentences are pathological and never arise in practice, therefore the issue can just be ignored. However even very simple examples of reasoning across contexts, applying for instance what McCarthy and Guha call *lifting rules*, exhibit the use of logical properties which subsume those required by Montague's result. Once these properties are available and since a paradoxical sentence can always be built in such systems, the whole logical formalism collapses. We will show how this happens in section 3.

A simple way out of paradoxes is to keep the object language separate from the metalanguage [Moore 77, Weyhrauch 80] and to build a hierarchy of languages [Konolige 82], each one being a meta-language for the previous, when nested beliefs are involved. Self reference

is not allowed and the construction of paradoxical sentences is blocked. However this forbids also non paradoxical self referential sentences like *John believes that he has a false belief* and mutually referential sentences like when one person says *I believe that we both believe true things* while his opponent denies it: *I do not believe so*. This lack of expressiveness may be considered a major drawback [Perlis 85, Davies 90] since self referential statements about truth, beliefs or knowledge arise naturally in common sense reasoning.

The complex machinery required by the layered approach also does not seem convenient for implementation within reasoning programs nor natural as a formalisation of common sense.

A more promising approach is the one suggested by Turner [Turner 89] and adopted also by the Interlingua working group of the DARPA Knowledge Sharing Initiative [Genesereth 92]. The solution advocated for the liar paradox is to build the extension of the truth predicate progressively in a layer-by-layer fashion like in [Kripke 75]. However the other modal operators are not covered and will have to be defined separately.

That we need a single theory where truth, propositional attitudes such as belief and knowledge, and situations can coexist is apparent if we envision computer agents able to plan and execute actions by taking into account their own beliefs and knowledge as well as the beliefs and knowledge of other agents.

Introducing a metalanguage is only the first step in this direction allowing one to express modalities as predicates: for each of them one still has to provide an axiomatization, verify the consistency and analyze the interactions among the modalities. This may lead to a proliferation of theories [Turner 89]. What we propose instead is to rely on the single notion of *entailment in viewpoint*, in terms of which modalities and several other notions can be defined. The solution for the paradoxes and the proof of consistency will apply to all these notions.

The *theory of viewpoints* that we propose is a reflective first order logic which amalgamates object and metalanguage by using *names* for each term and sentence of the language and which contains an axiomatization of provability and carefully formulated reflection rules which result in a non conservative but consistent extension. To develop its semantics we exploit some recent results [Gupta 82, Herzberger 82, Turner 89] on semantic theories of truth.

*Viewpoints* denote *sets of sentences* which represent the assumptions of a theory. A sentence of the form $in('A', vp)$, where $vp$ is a viewpoint expression, is interpreted as sentence $A$ is entailed by the assumptions denoted by $vp$.

In the theory of viewpoints modalities are expressed as entailment in a viewpoint: differences in the behaviour of modal operators are captured by additional axioms imposed on viewpoints, thus constraining or expanding the set of sentences that hold in specialised viewpoints [Simi 91]. Several concepts, even beyond the traditional modal operators, can be expressed through viewpoints: therefore viewpoints fulfill our goal of a unified framework in which all these notions can be handled and where one can reason about their interaction.

Viewpoints were conceived [Attardi 84] as a formalization of the notion of *context*, whose importance for knowledge representation has been stressed recently by several authors [McCarthy 87, McCarthy 93, Sowa 90, Guha 91]. Contexts should be considered as first class objects in the language so that they can serve as a frame of reference for sentences (which are relative to some context) and that it is possible to formulate rules which relate

one context to another (called lifting rules in [Guha 91]).

Several problems are to be addressed and critical choices are to be made to achieve a satisfactory theory of viewpoints: selecting appropriate model theoretic domains for interpretation to allow for self referential or mutually referential viewpoints; deciding between a two or three valued logic; having a single language or a hierarchy of languages; amalgamating object and metalevel or providing a fully reflective theory.

In each case we settled for what we considered the most advanced alternative: a two valued logic to maintain the naturalness and simplicity of classical logic; a single language instead of a hierarchy of languages; a fully reflective theory where mixed object/metalevel statements are allowed and can be proved.

The idea of viewpoints was introduced in [Attardi 84] and viewpoints were incorporated in the implementation of the Omega system [Attardi 86]. The basic theory of viewpoints was sketched in [Attardi 91] as a non conservative extension of first order predicate calculus, with the inclusion of an axiomatization of provability at the meta-level and of carefully formulated reflection rules. The consistency of the resulting theory was proved by establishing a correspondence with a syntactic variant of a modal logic, proved consistent by [Davies 90].

Such translation however did not provide sufficient insight into the real meaning of sentences relativised to viewpoints. The main goal of this paper is to develop an independent semantics for viewpoints, which is useful in better understanding how closely the semantics matches the proof theory. For providing a semantics to sentences of the form $\text{in}('A',\ vp)$, we introduce the notion of *contextual entailment*, as a way to express the fact that $A$ should hold in all interpretations which are coherent with the interpretation given to $vp$. The proof theory is sound with respect to this interpretation, and we obtain several hints on possible ways to strengthen it to get closer to completeness, which is nevertheless unachievable.

The paper is organised as follows: section 2 and 3 present the syntax and proof theory of viewpoints. Section 4 is devoted to a semantic account. Section 5 is a discussion on the intended use of viewpoints. Finally in the remaining sections we discuss extensions, implementation issues and draw some conclusions.

## 2 Syntax

The logic of viewpoints has the syntax of classical first order logic, extended with the sentences $\text{in}('A',\ vp)$, where $'A'$ is the name of a sentence and $vp$ can be:

1. a finite set of sentence names: $\{'A_1','A_2',...,'A_n'\}$

2. a viewpoint term, consisting of a viewpoint function and a list of terms. A viewpoint function of no argument is a viewpoint constant.

For each symbol $s$ of the language (either variable, function, predicate or viewpoint), we denote by $'s'$ its name. Names for terms and literals are obtained through the following definition:

$$'f(t_1, t_2, ..., t_n)' = \text{term}('f', 't_1', 't_2', ..., 't_n')$$

$$'p(t_1, t_2, ..., t_n)' = \text{literal}('p', 't_1', 't_2', ..., 't_n')$$

$$'A_1 \wedge A_2' = \mathsf{and}('A_1','A_2')$$

$$'A_1 \vee A_2' = \mathsf{or}('A_1','A_2')$$

$$'\neg A_1' = \mathsf{not}('A_1')$$

$$'\forall x.A' = \mathsf{forAll}('x','A')$$

$$'\mathsf{in}(t,\{t_1,t_2,...,t_n\})' = \mathsf{inTerm}('t',\mathsf{set}(t_1,t_2,...,t_n))$$

$$'\mathsf{in}(t,vp(t_1,t_2,...,t_n))' = \mathsf{inTerm}('t',\mathsf{vpTerm}('vp','t_1','t_2',...,'t_n'))$$

In the following we use the conventions that $A$, $B$ are metavariables for statements, $'A'$, $'B'$ are quoted statements while $vp, vp_1, vp_2$ are metavariables for viewpoints, i.e. expressions denoting sets of sentences. To simplify the notation, we will sometimes use $vp, vp_1, vp_2$ for sets of sentences or for the sentence which is the conjunction of the sentences in the set denoted by a viewpoint. It should be clear from the context which is the intended meaning for such meta-variables.

# 3 Proof theory

The proof system for a logical system to be used as the underpinning for reasoning programs, must be consistent and simple to implement. Decidability is out of question, since the expressive power of an amalgamated system approaches that of second order logic, and likewise for completeness.

In this section we concentrate in developing a consistent set of axioms and inference rules for viewpoints. Since the axiomatisation remains within first-order logic, well established theorem proving techniques remain applicable. Also, the constraints on the applicability of the reification inference rule are easy to fulfill: it is just necessary to keep track of whether or not reflection has been used in the course of a proof.

The proof systems of viewpoints consists of the axioms for classical logic, plus a suitable axiomatisation of provability and reflection rules which link what is provable in the object and meta level.

## 3.1 Axiomatisation of provability

The axiomatisation of provability exploited here is quite standard [Bowen 82, Attardi 91]. We consider an axiomatisation of first order logic and introduce a meta inference rule for each of the object level inference rules. For example, the following meta-rule:

$$\frac{\mathsf{in}('B', vp \cup \{'A'\})}{\mathsf{in}('A \Rightarrow B', vp)}$$

will correspond to the following rule of implication introduction:

$$\frac{vp \cup \{A\} \vdash B}{vp \vdash (A \Rightarrow B)}$$

The following axiom schemata are included, so that viewpoints will behave as logical theories:

$Ax1.$ $\mathsf{in}('A', \{...,'A',...\})$

$Ax2.$ $\mathsf{in}('A', vp)$, for any logical axiom $A$ and viewpoint $vp$.

## 3.2 Reflection rules

Two reflection rules provide the link between object and meta-level:

$$\frac{vp_1 \vdash \mathsf{in}('A', vp_2)}{vp_1 \cup vp_2 \vdash A} \qquad\qquad (\textit{reflection})$$

$$\frac{vp \vdash_C A}{\vdash \mathsf{in}('A', vp)} \qquad\qquad (\textit{reification})$$

The notation $\vdash_C$ stands for "classically derivable" or "derivable without using the reflection rules". The choice of this formulation of the reflection rules was discussed thoroughly in [Attardi 91].

The *reflection* rule is to be compared with weaker versions of *reflection* rules, such as the ones in [Bowen 82, Wehyrauch 80], which are formulated as follows:

$$\frac{\vdash_{MT} \mathsf{PR}('A')}{\vdash_T A} \qquad\qquad (\textit{conservative reflection})$$

where $\mathsf{PR}$ means "provable" and $T$ and $MT$ are the object and metalevel theory respectively. The important difference is that the precondition of our rule does not require that $\mathsf{in}('A', vp_2)$ be proved with no premises, but it is sufficient that such premises be assumed in the proof of $A$. Therefore, unlike the previous proposals, the theory resulting from the addition of the reflection rules is a non conservative extension: new theorems are derivable which where not provable in the combined object and metalevel theory.

For instance the following theorem is easily proved:

$$\mathsf{in}('A', vp) \Rightarrow (vp \Rightarrow A)$$

This sentence can in fact be regarded as an alternative formulation of the *reflection* rule, meaning that if $\mathsf{in}$ holds then material implication holds as well.

We claim that this is a desirable property for the formalisation of propositional attitudes, in particular when $\mathsf{in}$ is interpreted as belief. We will return to this point in section 5.

As a consequence of the stronger *reflection*, the restriction imposed on the *reification* rule becomes necessary in order to avoid falling in an inconsistent theory [Montague 63].

The implications of Montague's negative results about inconsistency of theories which have enough technical machinery to represent their own syntax and strong reflection rules have been largely discussed in the literature. In these theories a self referential formula $R$ can be constructed such that:

**Lemma 1** *In a logical system expressive enough to represent its own syntax, there exists a formula $R$ such that:* $\vdash R \Leftrightarrow \mathsf{P}('\neg R')$

where $\mathsf{P}('A')$ stands for any predicate on names of formulae such as $\mathsf{True}$, $\mathsf{Bel}$ or our own $\mathsf{in}('A', \{\})$. A similar self referential formula was constructed in [Bowen 82] for a different purpose: to show the incompleteness of their amalgamated object meta-level theory.

The self referential formula $R$ leads to inconsistency when the predicate $\mathsf{P}$ satisfies the properties expressed in the following theorem.

**Theorem 1 (Montague)** *Let $T$ be a theory such that:*

$(i)$  $\vdash \mathsf{P}('A') \Rightarrow A$

$(ii)$  $\vdash \mathsf{P}('\mathsf{P}('A') \Rightarrow A')$

$(iii)$  $\vdash \mathsf{P}('A')$, *if* $\vdash_C A$

$(iv)$  *If*  $\vdash \mathsf{P}('A \rightarrow B')$ *and* $\vdash \mathsf{P}('A')$ *then*  $\vdash \mathsf{P}('B')$

$(v)$  *$T$ includes enough machinery to represent its own syntax*

*then $T$ is inconsistent.*

Here $\vdash_C A$ means that the proof of $A$ is carried out without using any of the schemata $(i - iv)$.

As Montague has noted, all the above properties seem quite reasonable and it is hard to conceive a logic of syntactic modality which could dismiss any of them.

To see how easy it is to overlook the problem of consistency, let's consider the sketch of proof about the blocks world presented in [McCarthy 93]. In one step of that proof the fact that block $A$ is on block $B$ in situation $S_0$ is inferred for context $c$, written as:

$c : \quad on(A, B, S_0)$

from the fact that $\mathsf{ist}(c, on(A, B, S_0))$ holds in another context $c_0$, written as:

$c_0 : \quad \mathsf{ist}(c, on(A, B, S_0))$

where $\mathsf{ist}(c, p)$ means "$p$ is true in context $c$". This deductive step corresponds to the application of a quite strong reflection rule:

$$\frac{c_1 \vdash \mathsf{ist}(c_2, A)}{c_2 \vdash A}$$

Such rule leads directly to property $(i)$ of Montague:

$c_0 : \quad \mathsf{ist}(c_0, A) \Rightarrow A$

Unrestricted reification is also used in a later step of the same proof: therefore we obtain

$c_0 : \quad If \ \vdash A \ then \ \vdash \mathsf{ist}(c_0, A)$

$c_0 : \quad \mathsf{ist}(c_0, \mathsf{ist}(c_0, A) \Rightarrow A)$

which correspond to properties $(iii)$ and $(ii)$ of theorem 1. Property $(iv)$, equivalent to Modus Ponens within a context, is also exploited in the same proof. As a consequence, any context like $c_0$ is inconsistent.

Instead, in the proof theory for viewpoints, because of the restriction imposed on the reification rule, it is not possible to prove the equivalent of property $(ii)$, i.e.:

$$in('in('A', \{\}) \Rightarrow A', \{\})$$

which is essential and sufficient, together with the other properties of Montagues's theorem, to obtain the inconsistency result.

To reach the evidence of consistency, we had however to prove the following

**Theorem 2 (consistency)** *The logic of viewpoints is consistent.*

**Proof.** The full proof appears in [Attardi 91]. The result was obtained by establishing a correspondence with a syntactic variant of a modal logic, proved consistent in [Davies 90]. In essence, viewpoint formulae are translated into formulae of a consistent logic of truth: a sentence of the form $in('A', vp)$ is mapped into a sentence of the form $True('vp \Rightarrow A')$ and conversely $True('A')$ is translated into $in('A', \{\})$. These translations allow us to prove the equivalence of the two logics. $\square$

## 3.3 Additional axioms and properties

Due to the restriction on the *reification* rule, only classical proofs, which do not use reflection, can be transferred into an arbitrary viewpoint. This is formally expressed as:

**Proposition 1 (Proof in context)** *If* $in('A', vp)$ *and* $A \vdash_C B$ *then* $in('B', vp)$

Note that a weaker condition, like $A \vdash B$ or $in('B', \{'A'\})$ instead of $A \vdash_C B$ would not entail the same conclusion. It is however easy to show that a proof, provided it is classical, can be performed at any level of nesting.

The following sentence represents an intuitive property of viewpoints, saying that if something holds in a viewpoint, then this very fact itself holds in the same viewpoint:

$$in('A', vp) \Rightarrow in('in('A', vp)', vp) \qquad\qquad (positive\ introspection)$$

This property is usually referred as *positive introspection*, since if we interpret in as knowledge (or belief), it reads as: whatever one knows, one knows that he knows.

The truth of $in('A', vp)$ does not imply the existence of a classical proof, and therefore *reification* cannot be used for deriving the consequent of this sentence. This property is independent from the other axioms, and we choose therefore to add it explicitly to our axiomatisation.

The *reification* rule instead turns out to be derivable. This is due to the fact that a proof at the object level can be mirrored completely at the metalevel. Vice versa, if we carry out a proof at the metalevel without using any premises, the same conclusion could have been reached directly at the object level. The following theorem formalises these observations.

**Theorem 3** $\vdash in('A', vp) \quad iff \quad vp \vdash_C A$

This basic proof theory for viewpoints behaves reasonably well for the use that we have in mind. Later in section 5, we will give a sketch of how to use viewpoints to model beliefs, truth, knowledge and situations. Here we give an example of a proof involving three related viewpoints.

### 3.4 An example

We show here how to formalise a classic logical puzzle by means of viewpoints and then how to use the proof theory of viewpoints to solve it.

> *The members of the tribe of Voo always speak the truth, while the members of the tribe of Doo always speak the contrary. An explorer meets a Voo and a Doo, and has to figure out which is which.*

By denoting with $vp$ the common knowledge that both Voos and Doos have, we can write:

1. $Says(Voo, x) \Rightarrow \mathsf{in}(x, vp_V)$

2. $Says(Doo, x) \Rightarrow \mathsf{in}(\neg x, vp_D)$

where

3. $vp_V = vp \cup \{'Says(Doo, x) \Leftrightarrow \mathsf{in}(\neg x, vp)'\}$

4. $vp_D = vp \cup \{'Says(Voo, x) \Leftrightarrow \mathsf{in}(x, vp)'\}$

If the explorer asks to one native what the other would answer to a yes/no question, the answer $A$ he would get from a Voo would be false; in fact

5. $Says(Voo, 'Says(Doo, A)')$

6. $\quad \Rightarrow \mathsf{in}('Says(Doo, A)', vp_V)$ — MP(1, 5), proof in context

7. $\quad \Rightarrow \mathsf{in}('\mathsf{in}(\neg A, vp)', vp_V)$ — MP(2, 6), proof in context

8. $\quad \Rightarrow \mathsf{in}('vp \Rightarrow \neg A', vp_V)$ — MP(7, reflection), proof in context

9. $\quad \Rightarrow (vp_V \Rightarrow (vp \Rightarrow \neg A))$ — reflection(8)

10. $\quad \Rightarrow (vp \Rightarrow \neg A)$ — absorption(3)

Similarly the answer of a Doo would be false:

11. $Says(Doo, 'Says(Voo, A)')$

12. $\quad \Rightarrow \mathsf{in}('\neg Says(Voo, A)', vp_D)$

13. $\quad \Rightarrow \mathsf{in}('\neg\mathsf{in}(A, vp)', vp_D)$

14. $\quad \Rightarrow \mathsf{in}('\mathsf{in}(\neg A, vp)', vp_D)$

15. $\quad \Rightarrow \mathsf{in}('vp \Rightarrow \neg A', vp_D)$

16. $\quad \Rightarrow (vp_D \Rightarrow (vp \Rightarrow \neg A))$

17. $\quad \Rightarrow (vp \Rightarrow \neg A)$

Step 13 is justified by the choice of a yes/no question, for whose answer $A$ it is either $\mathsf{in}(A, vp)$ or $\mathsf{in}(\neg A, vp)$. Such is the case for instance for the answer *I am a Voo*. Therefore a suitable question to solve the puzzle could be: *What would your friend answer to the question: "Are you a Voo?"*.

# 4  Semantics

The main problem in providing a semantics to a reflective theory is to avoid the inconsistencies arising from sentences of the form $R \Leftrightarrow F('\neg R')$, which can be built by diagonalisation. Solutions proposed in the context of logics for truth, can be a source of inspiration. Kripke's account of truth [Kripke 75] is based on Kleene's three valued logic and on an iterative process which assigns truth values to larger and larger sets of sentences. In the final model all sentences are either true or false, satisfying our requirement for a binary logic. Unfortunately this model is too weak since not always $\mathsf{in}('A', \{\})$ when $A$ is a classical tautology like $B \vee \neg B$. A counterexample is $R \vee \neg R$, where $R$ is a paradoxical sentence.

We adopt therefore a different solution which builds upon the Gupta-Herzberger semantic theory [Gupta 82, Herzberger 82] along the lines indicated by Turner for a logic of truth [Turner 89].

An adequate semantics must also avoid to sanction dangerous formulae. For example in consistent reflective theories, a **T** axiom, i.e. an axiom such as our

$$\mathsf{in}('A', vp) \Rightarrow (vp \Rightarrow A)$$

is often allowed, but in this case the *reification* rule (called *necessitation* in the context of modal logics) must be restricted so that

$$\mathsf{in}('\mathsf{in}('A', vp) \Rightarrow (vp \Rightarrow A)', \{\})$$

is not derivable. Its derivation in fact would immediately lead to an inconsistent theory [Montague 63]. The same argument applies to the reified positive introspection.

Another problem we had to face is that we want to account for positive introspection, which is a useful principle for our purposes, without also having to admit the slightly different, stronger

$$\mathsf{in}('A', vp_1) \Rightarrow \mathsf{in}('\mathsf{in}('A', vp_1)', vp_2)$$

which is not desirable and in fact cannot be proved in our theory. As a consequence, an interpretation of $\mathsf{in}$ as straightforward entailment or derivability is not satisfactory. Our answer to this problem is an interpretation of $\mathsf{in}$ as *contextual entailment*, i.e. entailment restricted to a subclass of all models, to ensure that viewpoints at different levels of nesting are interpreted coherently. We want to interpret $\mathsf{in}(t_1, t_2)$ in a certain model $\mathcal{M}$, as entailment of $t_1$ from $t_2$, i.e. that whenever $t_2$ holds, then $t_1$ holds. However $\mathcal{M}$ establishes an interpretation for viewpoint functions some of which may appear in $t_1$; we must ensure that the same interpretation is used in evaluating $t_1$. Consider for instance:

$$\mathsf{in}('A', vp) \Rightarrow \mathsf{in}('\mathsf{in}('A', vp)', vp)$$

When examining the validity of the consequent, we are already restricted to models where $vp$ is a viewpoint which entails $A$. Therefore, when we consider whether $vp$ entails $\mathsf{in}('A', vp)$, we must carry over this restriction, and in fact we conclude that in all such models $A$ is entailed by $vp$.

We describe here our semantics. An interpretation structure $\mathcal{M}$ is a pair $\langle \mathcal{D}, \mathcal{I} \rangle$ defined as follows:

9

1. The domain of interpretation $\mathcal{D}$ contains the set of sentences $\mathcal{S}$ of the language.

2. $\mathcal{I}$ associates an $n$-ary function $f_{\mathcal{I}} : \mathcal{D}^n \to \mathcal{D}$, to each $n$-ary function symbol $f$ of the language, except the syntax constructors term, literal, and, or, not, forAll, set, vpTerm and inTerm, whose interpretation is fixed.

3. $\mathcal{I}$ associates to each $n$-ary predicate symbol $p$ of the language, an $n$-ary predicate $p_{\mathcal{I}} \subseteq \mathcal{D}^n$.

4. $\mathcal{I}$ associates to each $n$-ary viewpoint function $vp$, a function $vp_{\mathcal{I}} \to 2^{\mathcal{S}}$

Notice that since a domain $\mathcal{D}$ must contain sentences, and sentences may represent viewpoints, which are in turn sets of sentences, and finally a viewpoint can refer to its own sentences, as suitable domains one might need Scott's lambda calculus models or Aczel's non well founded sets [Aczel 88].

Terms are interpreted with respect to an assignment function $g$, which assigns elements of $\mathcal{D}$ to variables. The interpretation of terms is classical except for naming terms whose interpretation is the term or sentence they name. In detail:

$[\![x]\!]_g = g(x)$

$[\![f(t_1, t_2, ..., t_n)]\!]_g = f_{\mathcal{I}}([\![t_1]\!]_g, [\![t_2]\!]_g, ..., [\![t_n]\!]_g)$, where $f$ is not a syntax constructor.

$[\![{}'x']\!]_g = x$

$[\![\text{literal}(t, t_1, t_2, ..., t_n)]\!]_g = [\![t]\!]_g([\![t_1]\!]_g, [\![t_2]\!]_g, ..., [\![t_n]\!]_g)$

$[\![\text{term}(t, t_1, t_2, ..., t_n)]\!]_g = [\![t]\!]_g([\![t_1]\!]_g, [\![t_2]\!]_g, ..., [\![t_n]\!]_g)$

$[\![\text{vpTerm}(t, t_1, t_2, ..., t_n)]\!]_g = [\![t]\!]_g([\![t_1]\!]_g, [\![t_2]\!]_g, ..., [\![t_n]\!]_g)$

$[\![\text{and}(t_1, t_2)]\!]_g = [\![t_1]\!]_g \wedge [\![t_2]\!]_g$

$[\![\text{or}(t_1, t_2)]\!]_g = [\![t_1]\!]_g \vee [\![t_2]\!]_g$

$[\![\text{not}(t)]\!]_g = \neg [\![t]\!]_g$

$[\![\text{forAll}({}'x', t)]\!]_g = \forall x. [\![t]\!]_g$

$[\![\text{inTerm}(t_1, t_2)]\!] = \text{in}([\![t_1]\!]_g, [\![t_2]\!]_g)$

Moreover for sets of sentence names and viewpoint terms (viewpoint constants as a special case) we have:

$[\![\text{set}(t_1, t_2, ..., t_n)]\!]_g = \{[\![t_1]\!]_g, [\![t_2]\!]_g, ..., [\![t_n]\!]_g\}$

$[\![vp(t_1, t_2, ..., t_n)]\!]_g = vp_{\mathcal{I}}([\![t_1]\!]_g, [\![t_2]\!]_g, ..., [\![t_n]\!]_g)$

In short we could write, for any term t: without variables

$[\![{}'t']\!] = t$

10

For each closed formula $A$, $\models_{\mathcal{M},g} A$ means that $A$ is true in the interpretation $\mathcal{M}$ with assignment function $g$. The notation $g[i/x]$ represents the assignment function identical to $g$ except that $x$ is bound to $i$.

For the interpretation of sentences we proceed by defining the notion of true at level $n$, where the level corresponds to the level of nesting of in sentences. A sentence with a nesting level of $n$ will receive its definitive truth value at level $n$; while paradoxical sentences, to which no finite level of nesting can be assigned, will keep oscillating periodically between true and false.

At level 0, all sentences are considered false: not $\models_{\mathcal{M},g}^{0} A$.

For any $n > 0$:

$$\models_{\mathcal{M},g}^{n} p(t_1, t_2, ..., t_n) \quad \text{iff} \quad \langle [\![t_1]\!]_g, [\![t_2]\!]_g, ..., [\![t_n]\!]_g \rangle \in p_{\mathcal{I}}$$

$$\models_{\mathcal{M},g}^{n} (\neg A) \quad \text{iff} \quad \text{not } \models_{\mathcal{M},g}^{n} A$$

$$\models_{\mathcal{M},g}^{n} (A \wedge B) \quad \text{iff} \quad \models_{\mathcal{M},g}^{n} A \text{ and } \models_{\mathcal{M},g}^{n} B$$

$$\models_{\mathcal{M},g}^{n} (A \vee B) \quad \text{iff} \quad \models_{\mathcal{M},g}^{n} A \text{ or } \models_{\mathcal{M},g}^{n} B$$

$$\models_{\mathcal{M},g}^{n} (\forall x.A) \quad \text{iff} \quad \text{for all } d \text{ in } \mathcal{D}, \models_{\mathcal{M},g[d/x]}^{n} A$$

The truth of in sentences at level $n$ is defined on the basis of the truth at the level $n-1$. The interpretation of in is *contextual entailment*, i.e. entailment restricted to a subclass of all models, to ensure that viewpoints at different levels of nesting are interpreted coherently:

$$\models_{\mathcal{M},g}^{n} \mathsf{in}(t_1, t_2) \quad \text{iff} \quad \text{for all } \mathcal{N} \in \mathcal{M}[t_2], \text{ not } \models_{\mathcal{N},g}^{n-1} [\![t_2]\!]_g \text{ or } \models_{\mathcal{N},g}^{n-1} [\![t_1]\!]_g$$

where $\mathcal{M}[t]$ is the class of models that coincide with $\mathcal{M}$ on all the viewpoint functions appearing in $t$, i.e. $\mathcal{M}[t] = \{\langle \mathcal{D}, \mathcal{I}' \rangle |$ for each viewpoint function $f$ appearing in $t$, $f_{I'} = f_I\}$. Note the difference in semantics between implication and in operators: in might be interpreted as "necessary implication".

In most cases, this process of revision stabilises in the sense that the truth value of sentences settles to either true or false from a certain level onward. There are however paradoxical sentences, like the counterpart of the liar sentence, which do not stabilise at any finite level of revision but continue to oscillate between different truth values. To such sentences false will be assigned as ultimate truth value.

A sentence is said to be *stably true* accordingly to whether or not the revision process stabilises:

$$\models_{\mathcal{M},g}^{*} A \quad \text{iff} \quad \exists k. \forall n > k. \models_{\mathcal{M},g}^{n} A$$

Stable truth is however too coarse, since all sentences which do not stabilise are assigned to false: therefore for any such sentence $R$ both $R$ and $\neg R$ would be false. One further step is necessary to separate among the sentences which do not stabilise.

The notion of truth hinges on that of stable truth for in sentences:

$$\models_{\mathcal{M},g} p(t_1, t_2, ..., t_n) \quad \text{iff} \quad \langle [\![t_1]\!]_g, [\![t_2]\!]_g, ..., [\![t_n]\!]_g \rangle \in p_{\mathcal{I}}$$

$$\models_{\mathcal{M},g} (\neg A) \quad \text{iff} \quad \text{not } \models_{\mathcal{M},g} A$$

$$\models_{\mathcal{M},g} (A \wedge B) \quad \text{iff} \quad \models_{\mathcal{M},g} A \text{ and } \models_{\mathcal{M},g} B$$

$$\models_{\mathcal{M},g} (A \vee B) \quad \text{iff} \quad \models_{\mathcal{M},g} A \text{ or } \models_{\mathcal{M},g} B$$

$$\models_{\mathcal{M},g} (\forall x.A) \quad \text{iff} \quad \text{for all } d \text{ in } \mathcal{D}, \models_{\mathcal{M},g[d/x]} A$$

$$\models_{\mathcal{M},g} \mathsf{in}(t_1, t_2) \quad \text{iff} \quad \models^*_{\mathcal{M},g} \mathsf{in}(t_1, t_2)$$

**Note.** The Gupta-Herzberger semantics is more complex and requires transfinite induction to distinguish different degrees of instability for formulae. For instance if we have two formulae $A$ and $B$ which oscillate out of phase, according to their semantics the formula $(A \vee B)$ will be stable, and in fact true, while in our semantics it will be false. For example consider $A = $ *"This sentence is false"*, $B = $ *"It is false that what I say is true"*. We could make it true as well by the following clause:

$$\models_{\mathcal{M},g} (A \vee B) \quad \text{iff} \quad \models_{\mathcal{M},g} A \text{ or } \models_{\mathcal{M},g} B \text{ or } \models^*_{\mathcal{M},g} (A \vee B)$$

However we will not be able to prove it true with our proof theory, so it is not clear whether it is worth to make this change.

Validity is defined as usual:

$$\models A \quad \text{iff} \quad \forall \mathcal{M}, g. \models_{\mathcal{M},g} A$$

$vp \models A$ means that $A$ is true in every model of $vp$.

We can now verify that our semantics has the required properties.

**Note.** The *law of excluded middle* holds both at the object level:

$$A \vee \neg A$$

and at the metalevel:

$$\mathsf{in}('A \vee \neg A', vp)$$

However this does not imply that $\mathsf{in}('A', vp) \vee \mathsf{in}('\neg A', vp)$, which would be undesirable since it would sanction the completeness of any viewpoint.

Consider the formula $R$, such that $R \Leftrightarrow \mathsf{in}('\neg R', \{\})$, which corresponds to the Liar sentence. The paradox is avoided since the Liar sentence is false in every model:

**Proposition 2** $\models \neg R$.

**Proof.** For every $n > 0$, $\models^n_{\mathcal{M}} \neg R \quad \text{iff} \quad \models^{n+1}_{\mathcal{M}} R$:

$$\models^n_{\mathcal{M}} \neg R$$

$$\text{iff} \quad \models^n_{\mathcal{N}} \mathsf{in}(\neg R, \{\})$$

$$\text{iff} \quad \text{not for all } \mathcal{N}. \models^{n-1}_{\mathcal{N}} \neg R$$

$$\text{iff} \quad \text{not} \models^n_{\mathcal{K}} \mathsf{in}(\neg R, \{\}) \qquad \qquad \qquad \text{(for an arbitrary } \mathcal{K})$$

iff    for all $\mathcal{K}$ . not $\models^n_\mathcal{K} R$                                                   (by definition of $R$)

iff    for all $\mathcal{K}$. $\models^n_\mathcal{K} \neg R$

iff    $\models^{n+1}_\mathcal{N} \mathsf{in}('\neg R', \{\})$

iff    $\models^{n+1}_\mathcal{N} R$

<div align="right">□</div>

**Proposition 3** $\models^*_\mathcal{M} A$ *implies* $\models_\mathcal{M} A$ *but not vice versa.*

**Proof.** As a counterexample consider the case $\neg R$.                                 □

**Lemma 2** $\mathsf{in}('A', \{'B'\}) \Rightarrow (A \Rightarrow B)$ *is valid.*

**Proof.**

$\forall \mathcal{M}. \models_\mathcal{M} \mathsf{in}('A', \{'B'\}) \Rightarrow (A \Rightarrow B)$

not $\models_\mathcal{M} \mathsf{in}('A', \{'B'\})$ or $\models_\mathcal{M} (B \Rightarrow A)$

Assume $\models_\mathcal{M} \mathsf{in}('A', \{'B'\})$, then by definition:

$\models^*_\mathcal{M} \mathsf{in}('A', \{'B'\})$

for all $\mathcal{N} \in \mathcal{M}[A]$ . not $\models^n_\mathcal{N} A$ or $\models^n_\mathcal{N} A$                               (for $n > k$)

Since $\mathcal{M} \in \mathcal{M}[B]$, not $\models^n_\mathcal{M} B$ or $\models^n_\mathcal{M} A$:

$\models^n_\mathcal{M} B \Rightarrow A$                                                       (for $n > k$)

$\models_\mathcal{M} B \Rightarrow A$

<div align="right">□</div>

**Corollary 1** $(\mathsf{in}('A', \{\}) \Rightarrow A)$ *is valid.*

**Lemma 3** $\mathsf{in}('\mathsf{in}('A', \{\}) \Rightarrow A', \{\})$ *is not valid.*

**Proof.** (*Sketch*) It is not difficult to show that $(\mathsf{in}('R', \{\}) \Rightarrow R)$ does not stabilise.     □

**Corollary 2** $\mathsf{in}('\mathsf{in}('A', \{'B'\}) \Rightarrow (B \Rightarrow A)', vp)$ *is not valid.*

**Lemma 4 (Positive introspection)** $\mathsf{in}('A', vp) \Rightarrow \mathsf{in}('\mathsf{in}(A, vp)', vp)$, *is valid.*

**Proof.** (*Sketch*) The interpretation of $\mathsf{in}$ as contextual entailment, ensures a coherent interpretation of the inner and outer occurrence of $vp$.     □

**Note.** Like lemma 2, lemma 4 is valid but cannot be used in *reification*, since it cannot be derived classically, and in fact $\mathsf{in}('\mathsf{in}('A', vp) \Rightarrow \mathsf{in}('\mathsf{in}('A', vp)', vp)', \{\})$ is not valid.

**Lemma 5** $\mathsf{in}('A', vp) \Rightarrow \mathsf{in}('\mathsf{in}('A', vp)', vp_2)$ *is not valid.*

Now that we have a satisfactory semantic account for our theory of viewpoints, we can investigate how faithful is our axiomatisation to the proof theory. The first obvious result is soundness.

**Theorem 4 (Soundness)** *If $\vdash A$ then $\models A$.*

We turn our attention to the reflection rules, to discover that they are not yet ideal.

To stay safe of paradoxes, we have chosen a rule for *reification* which is fairly weak. Even though we could never achieve completeness for our logic, we might want to look for a stronger version of *reification*. What the semantics suggests is that $\mathsf{in}('A', vp)$ is valid whenever $A$ is grounded and it is entailed by $vp$.

**Definition 1** *A formula $A$ is* grounded *iff* $\models^*_\mathcal{M} A$ *or* $\models^*_\mathcal{M} \neg A$.

The *reification* rule could be formulated as follows:

$$\frac{vp \vdash A, Grounded(A)}{\vdash \mathsf{in}('A', vp)}$$

A similarly restricted necessitation rule is proposed by Turner and Davies [Davies 90, Turner 89] in connection with logics for truth.

This formulation however has the disadvantage that grounded sentences cannot be characterised syntactically, violating our requirement of a mechanisable proof system.

## 5 Use of viewpoints

We briefly report here on the use of viewpoints as a unifying framework for the representation of beliefs, truth, knowledge and situations: all these notions can be expressed as provability in a viewpoint, as was shown in [Simi 91]. Differences in the behaviour of modal operators are captured by additional axioms imposed on viewpoints, thus constraining or expanding the set of sentences that can be proved in specialised viewpoints.

This approach proves to be a valid alternative to other proposals were separate groups of axioms are to be provided for each modality, together with axioms relating different modalities with each other. The proof of consistency needs to be redone any time a new modality is added. The proof of consistency of the basic theory of viewpoints is done once for all. The worst that it can happen is that if a viewpoint is too much constrained by additional axioms a model for it will lack but the basic theory will remain consistent. We will make this more clear when discussing the formalisation of truth.

In the rest of the paper we simplify the notation further by getting rid of quotes, which become unessential details.

## 5.1 Beliefs

The beliefs of an agent $g$ is rendered directly by means of in sentences, by using $vp(g)$ as the viewpoint corresponding to the set of assumptions of the agent. More precisely:

$$\mathsf{Bel}(g, A) = \mathsf{in}(A, vp(g)) \qquad\qquad\qquad (\textit{definition of beliefs})$$

We can postulate, if we wish, a kind of internal coherence for agent beliefs, with an additional axiom of consistency, which is not required by the basic theory.

$$\neg\mathsf{in}(\mathsf{false}, vp) \qquad\qquad\qquad (\textit{consistency of viewpoints})$$

With this additional axiom all the properties usually ascribed to beliefs can be derived. For example, the axiom **K** of modal logics, i.e.

$$\mathsf{in}(A \Rightarrow B, vp) \Rightarrow (\mathsf{in}(A, vp) \Rightarrow \mathsf{in}(B, vp))$$

is a property of in which corresponds to the metalevel version of *modus ponens*.

The axiom **D** of modal logics, i.e.

$$\mathsf{in}(A, vp) \Rightarrow \neg\mathsf{in}(\neg A, vp)$$

postulates a kind of coherence for an agent's beliefs and can be derived by using the consistency axiom. Without it, a weaker property would hold, i.e.

$$vp \Rightarrow (\mathsf{in}(A, vp) \Rightarrow \neg\mathsf{in}(\neg A, vp))$$

The Barcan formula, i.e.

$$\forall x \quad \mathsf{in}(A, vp) \Rightarrow \mathsf{in}(\forall x A, vp)$$

can easily be proved with the rule of $\forall$-elimination and the meta-rule for $\forall$-introduction.

Moreover the property of positive introspection, which is usually considered a valid property for belief, is an axiom of the basic theory of viewpoints and our reification rule corresponds to a restricted rule of necessitation.

Finally, the rule of *reflection*, i.e.

$$\mathsf{in}(A, vp) \Rightarrow (vp \Rightarrow A)$$

is perfectly reasonable and even useful under the interpretation of in as belief, even though it roughly corresponds to a knowledge axiom. As stated, it corresponds to the following property of beliefs: *if an agent holds belief A, then either some of the agent's assumptions are false or belief A is true.* This property was exploited also in the solution to the puzzle presented in section 3, where conclusions were drawn based on an agent's trust or mistrust.

## 5.2 Truth

Truth is rendered as provability in a special theory, that we call Real World (RW). Ideally, everything that it is true should be derivable in this theory. We are therefore assuming the following definition for truth:

$\mathsf{True}(A) = \mathsf{in}(A, \mathsf{RW})$                                                                                      (*definition of truth*)

We might be tempted to give a characterisation of RW by assuming the equivalence $\mathsf{in}('A', \mathsf{RW}) \Leftrightarrow A$ as for example in [Konolige 82]. It is known however that this leads immediately to the Tarski's liar paradox. Therefore weaker properties should be explored, that our intuition feels appropriate for a theory of the real world, namely:

*consistency*: the real world should be non contradictory;

*completeness*: something is true or it is not;

*veridicality*: the set of assumptions of the real world are true.

This amounts to imposing the following properties as axioms, whatever set of assumptions RW is chosen as a description of the real world:

$\neg\mathsf{in}(false, \mathsf{RW})$                                                                                      (*consistency of* RW)

$\neg\mathsf{in}(A, \mathsf{RW}) \Rightarrow \mathsf{in}(\neg A, \mathsf{RW})$                                                            (*completeness of* RW)

RW                                                                                      (*veridicality of* RW)

The above properties are not minimal; in fact veridicality implies consistency. Moreover it can be shown that veridicality and completeness together are too powerful, since $A \Rightarrow \mathsf{in}(A, \mathsf{RW})$ can be proved using them: the liar paradox springs up again.

Among the two possible ways out, to give up veridicality or completeness, we choose to go for the second alternative because completeness is problematic in itself. In fact well known results show that no axiomatisation of provability can be complete (see for example [Bowen 82]); this concretely means that for any $vp$, a sentence $J$ can be constructed such that it says of itself that it is unprovable in $vp$, i.e.

$J \Leftrightarrow \neg\mathsf{in}(J, vp)$

can be proved. It follows that, postulating completeness of a viewpoint, precludes the existence of a model for that viewpoint. In fact $\neg vp$ can be derived both from the assumption $\mathsf{in}(J, \ vp)$ and from the assumption $\neg\mathsf{in}(J, \ vp)$. This can also be alternatively stated as "no truthful set of assumptions can be complete" or "no complete set of assumptions can be consistent".

This negative result suggest to investigate weaker forms of completeness. One restriction that comes naturally to mind is to impose some kind of *closed world assumption*, that is to insist on completeness only for ground atomic formulas, not including in statements. We will assume the existence of a predicate $SAF$ (Simple Atomic Formulas) on the names of sentences to test this condition. The restricted completeness is:

$SAF(A) \wedge \neg\mathsf{in}(A, \mathsf{RW}) \Rightarrow \mathsf{in}(\neg A, \mathsf{RW})$                                        (*weak completeness of* RW)

16

Now the double implication, $A \Leftrightarrow \mathsf{in}(A, \mathsf{RW})$, does not hold in general, and this is enough to prevent the liar paradox. Therefore in addition to positive introspection and consistency, we will have the above weak completeness and veridicality.

$$\mathsf{RW} \hspace{11cm} (\textit{veridicality of}\ \mathsf{RW})$$

We regard this solution as the minimal set of properties for $\mathsf{RW}$ allowing to characterise this special viewpoint in such a way that paradoxes are avoided. It is worth noting that the axioms imposed on $\mathsf{RW}$, and those imposed on viewpoints for beliefs, are to be considered logical axioms of an extended theory of viewpoints and as such, in virtue of $Ax2$, they are included in any viewpoint. Therefore, for example, since veridicality is a logical axiom, $\mathsf{in}(\mathsf{RW},\ vp)$ holds for any $vp$. However, the fact that any viewpoint is aware that $\mathsf{RW}$ holds, does not imply that any viewpoint is aware of all the true facts. In fact, the property:

$$\mathsf{in}(A, \mathsf{RW}) \Leftrightarrow \mathsf{in}(A, vp)$$

cannot be proved. This is due to the restriction on reification and the consequent restriction on the "proof in context" rule. When truth is used in connection with beliefs, we have:

$$\mathsf{in}(\mathsf{in}(A, vp), \mathsf{RW}) \Leftrightarrow \mathsf{in}(A, vp)$$

However the following cannot be derived, due again to the restriction on the reification rule:

$$\mathsf{in}(\mathsf{in}(A, \mathsf{RW}), vp) \Leftrightarrow \mathsf{in}(A, vp)$$

## 5.3  Knowledge

One way to model knowledge is as a true belief:

$$\mathsf{K}(g, A) = \mathsf{Bel}(g, A) \wedge \mathsf{True}(A) = \mathsf{in}(A, vp(g)) \wedge \mathsf{in}(A, \mathsf{RW}) \hspace{1cm} (\textit{definition of knowledge})$$

All the properties usually ascribed to knowledge can then be derived

$$\mathsf{K}(g, A) \Rightarrow A \hspace{9cm} (\textit{knowledge axiom})$$

which immediately follows from $\mathsf{in}(A, \mathsf{RW})$, by means of reflection and veridicality.

## 5.4  Situations

Similarly to what was done for beliefs, we associate to each situation $s$, a set of basic facts which define the situation, forming a viewpoint $vp(s)$, and define what holds in a situation as provability in such viewpoint.

$$\mathsf{Hold}(A, s) = \mathsf{in}(A, vp(s)) \hspace{4cm} (\textit{definition of "holding" in a situation})$$

Like for being true, holding in a situation is intuitively different from provable in a theory. Stronger properties should apply, like the fact that a sentence holds or does not

hold in a situation, completeness again. But, as discussed before, a restricted form of completeness is only possible. One can argue about consistency of situations: after all it is possible to hypothesise situations which are incoherent and have no hope of being real. Nevertheless we will assume consistency, while veridicality is definitely not appropriate for situations.

We will therefore characterise situations with the following axioms:

$$SAF(A) \wedge \neg\mathsf{in}(A, vp(s)) \Rightarrow \mathsf{in}(\neg A, vp(s)) \qquad (\textit{weak completeness of situations})$$

$$\neg\mathsf{in}(false, vp(s)) \qquad (\textit{consistency of situations})$$

The fact that veridicality does not apply, prevents the proof of

$$\mathsf{in}(A, s) \Rightarrow A$$

which would not be appropriate for situations.

In conclusion, the modalities that we have discussed are modeled as viewpoints with increasingly larger set of axioms, as summarized in this table:

| Beliefs | Situations | Truth |
|---|---|---|
| *Positive introspection* <br> *Consistency* | *Positive introspection* <br> *Consistency* <br> *Weak Completeness* | *Positive introspection* <br> *Consistency* <br> *Weak Completeness* <br> *Veridicality* |

An example combining all these notions was presented in [Simi 91]. We believe that other more esoteric propositional attitudes could be similarly defined.

## 5.5  McCarthy's Contexts

Our notion of viewpoint has much in common with recent proposals towards the formalisation of contexts [Guha 91, Shoham 91, McCarthy 93]. In fact the formalisation of some notion of context was also among our original motivations [Attardi 84].

A naive way of representing contexts with viewpoints would be to map them into viewpoints and use $\mathsf{in}$ to represent $\mathsf{ist}$:

$$\mathsf{ist}(c, p) = \mathsf{in}(p, c)$$

This translation would not capture however some aspects of McCarthy's contexts, like the fact that the meaning of a term may be different in different contexts. A simple technical way to mimic this feature would be to distinguish the symbols in various contexts by adding an index denoting their context. For instance, if we wanted to distinguish the meaning of *at* in context $c_1$

$$\mathsf{in}(at(jmc, Stanford), c_1)$$

where it means "works at", from its meaning in context $c_2$

$$\mathsf{in}(at(jmc, Stanford), c_2)$$

18

where it means "is present at", one could translate the first sentence into:

$$\mathsf{in}(at_{c_1}(jmc, Stanford), c_1)$$

It is not clear whether it is always necessary or appropriate to add such indexing to symbols. One might argue for instance that the difference in meaning between the two occurrences of *at* could be just as well captured by different assertions present in the two viewpoints $c_1$ and $c_2$. Different conclusions from these assertions could be drawn which would account for the difference in meaning in the two contexts.

One may question whether, for the notion of context envisaged in [McCarthy 93], the generality and the expressive power of reflexive viewpoints are not needed.

But, since contexts can be used for representing epistemic attitudes (see for example Shoham [Shoham 91]), self referential statements like *John believes that he has a false belief* should be accounted for:

$$\mathsf{in}(\exists A.\ \mathsf{in}(A \wedge \mathsf{False}(A), vp(John)), vp(John))$$

and it seems very unnatural to disallow self referential sentences just because they might cause trouble.

# 6  Exploiting the taxonomy of viewpoints

Given that viewpoints are sets, it is quite natural to think of exploiting the algebra of sets and to allow union, intersection and complement operations on viewpoints. Several interesting properties follow in an obvious way.

A very natural a subsumption relation among viewpoints is induced by the subset relation:

$$vp_1\ \mathsf{subsumes}\ vp_2 \equiv vp_1 \subseteq vp_2$$

The following property of monotonicity of viewpoints would then ensue:

$$vp_1\ \mathsf{subsumes}\ vp_2 \Leftrightarrow \mathsf{in}(A, vp_2) \Rightarrow \mathsf{in}(A, vp_1) \qquad (\textit{monotonicity of viewpoints})$$

Viewpoints, so defined, constitute a complete lattice whose top is the empty viewpoint, where only logical truth hold (we called it the *tautological viewpoint*), and whose bottom is the *contradictory viewpoint*, where anything holds.

An algebra of viewpoints along these lines was proposed together with the first version of viewpoints presented in [Attardi 84] and exploited in the implementation of the Omega system.

# 7  Implementation of viewpoints

The Omega system [Attardi 86] provides a viewpoints mechanism. Each viewpoint may represent a different theory and may be used to express different knowledge bases, or different situations, or different logic theories or to perform hypothetical reasoning. Viewpoints can be arranged in a hierarchy, where a viewpoint inherits from another by including all the

19

sentences that belong to it. Thereafter a viewpoint will inherit all the logical consequences from its ancestor viewpoints. This is useful for instance to create viewpoints describing a basic theory (e.g. natural deduction), from which more specific viewpoints can be created by adding new statements.

Operations are available to enter and exit a viewpoint, to create a viewpoint combining previous viewpoints, to assert a fact in a specific viewpoint.

The implementation uses bit masks to represent viewpoints, so that operations such as determining whether a statement belongs to a viewpoint or merging two viewpoints for combining knowledge from different sources, can be performed efficiently. The metalevel capability is used in the Omega system to associate deductive strategies to statements: since these can differ between viewpoints, the effect is to have different deductive capabilities in each viewpoint.

# 8    Conclusions

We have explored the idea that through the notion of viewpoint a variety of other concepts can be expressed, thereby providing a unifying framework for dealing with beliefs, knowledge and situations. In developing a theory of viewpoints several hard issues had to be dealt, such as those connected to the presence of paradoxical sentences which inevitably arise in reflective theories. The semantic account for viewpoints that we provided is an essential requirement for further developing this approach. The solution that we presented is simple, elegant and powerful compared to previous proposals for syntactic treatment of modalities. Since the semantics clearly identifies the sentences which lead to paradoxes, it is now possible to study further the reflection rules to determine conditions under which they can be strengthened.

The formalisation and the proof theory presented here can be adapted to the notion of context as conceived by McCarthy. We on purpose restrained to do so, since we wanted to keep our theory as simple as possible, without overloading it with features that might obscure its essence. For instance, McCarthy allows his contexts to have different vocabularies. While there are good motivations for this, it is not clear that this is an essential feature of the notion of context, since there are several ways to achieve the same effect. On the contrary, by sticking to a single vocabulary, we could develop a proof system which need not rely heavily on default reasoning.

# 9    References

Aczel, P., *Non-well-founded sets*, CSLI lecture notes, **12**, Center for the Study of Language and Information, Stanford, CA, 1988.

Attardi G., M. Simi, Metalanguage and reasoning across viewpoints, in *ECAI84: Advances in Artificial Intelligence*, T. O'Shea (ed.), Elsevier Science Publishers, Amsterdam, 1984.

Attardi G., M. Simi, A Description Oriented Logic for Building Knowledge Bases, *Proceedings of the IEEE*, **74**(10), October 1986, 1335-1344.

Attardi G., M. Simi, Reflections about reflection, in Allen, J. A., Fikes, R., and Sandewall, E. (eds.) *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*. Morgan Kaufmann, San Mateo, CA, 1991.

Bowen K.A., R.A. Kowalski, Amalgamating language and metalanguage in logic programming, in *Logic Programming*, K. Clark and S. Tarnlund (eds.), Academic Press, 1982, 153-172.

Davies N., Towards a first order theory of reasoning agents, *Proc. of 9th European Conference on Artificial Intelligence*, Stockholm, 1990, 195-200.

Genesereth M.R., R. Fikes, Knowledge Interchange Format Version 2.2 Reference manual, LG Report Logic-90-4, CS Department, Stanford, 1992.

Guha R. V., Contexts: a formalization and some applications, MCC Tech Rep ACT-CYC-42391, Nov. 1991.

Gupta A., Truth and paradox, *Journal of Philosophical Logic*, **11**, 1982, 1-60.

Herzberger H., Notes on naive semantics, *Journal of Philosophical Logic*, **11**, 1982, 61-102.

Konolige K., A first order formalization of knowledge and action for a multiagent planning system, *Machine Intelligence* 10, 1982.

Kripke S., Outline of a theory of truth, *Journal of Philosophy*, **22**(1), 1975, 690-716.

McCarthy J., First order theories of individual concepts and propositions, *Machine Intelligence* 9, 1979, 129-147.

McCarthy J., Generality in Artificial Intelligence, *Communications of the ACM*, **30**(12), 1987, 1030-1035.

McCarthy J., Notes on Formalizing Context, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 1993.

Montague R., Syntactical treatment of modalities, with corollaries on reflexion principles and finite axiomatizability, *Acta Philosoph. Fennica*, **16**, 1963, 153-167.

Moore R. C., Reasoning about knowledge and action, *Proc. of IJCAI77*, Cambridge, MA, 1977, 223-227.

Perlis D., Languages with self-reference I: foundations, *Artificial Intelligence*, **25**, 1985, 301-322.

Perlis D., Language with self reference II: knowledge, belief and modality, *Artificial Intelligence*, (34),2, 1988, 179-212.

Shoham, Y., Varieties of contexts, in Vladimir Lifschitz (ed.), *Artificial Intelligence and Mathematical Theory of Computation: papers in honor of John McCarthy*, Academic Press, 1991.

Simi M., Viewpoints subsume belief, truth and situations, in *Trends in Artificial Intelligence*, Proc. of 2nd Congress of the Italian Association for Artificial Intelligence, Ardizzone, Gaglio, Sorbello (Eds), *Lecture Notes in Artificial Intelligence* 549, Springer Verlag, 1991, 38-47.

Sowa J.F., Finding structure in knowledge soup, *Proc. Info Japan 90 - Information Processing Society of Japan*, 1990, 245-252.

Turner R., *Truth and modality for knowledge representation*, Pitman Press, London, 1989.

Weyhrauch R.W., Prolegomena to a theory of mechanized formal reasoning, *Artificial Intelligence*, **13** (1,2), 1980, 133-170.