Limited Logical Belief Analysis

Antonio Moreno
Departament d'Enginyeria Informàtica
Escola Tècnica Superior d'Enginyeria
Universitat Rovira i Virgili
Carretera de Salou, s/n. 43006-Tarragona, Spain
e-mail: amoreno@etse.urv.es
Phone: 34-77-559681
Fax: 34-77-559699

Abstract

The process of *rational inquiry* can be defined as the evolution of the beliefs of a rational agent as a consequence of its internal inference procedures and its interaction with the environment. These beliefs can be modelled in a formal way using *doxastic logics*. The *possible worlds model* and its associated *Kripke semantics* provide an intuitive semantics for these logics, but they seem to commit us to model agents that are *logically omniscient* and *perfect reasoners*. These problems can be avoided with a syntactic view of possible worlds, defining them as arbitrary sets of sentences in a propositional belief logic. In this article this syntactic view of possible worlds is taken, and a dynamic analysis of the agent's beliefs is suggested in order to model the process of *rational inquiry* in which the agent is permanently engaged. One component of this analysis, the *logical* one, is summarily described. This dimension of analysis is performed using a modified version of the *analytic tableaux method*, and it models the evolution of the beliefs due to the agent's inference power. It is shown how non-perfect reasoning is achieved in two ways: on one hand, the agent's deductive abilities can be controlled by restricting the tautologies that it is allowed to use in the course of this logical analysis; on the other hand, the agent is not obliged to perform an exhaustive analysis of the initial tableau[1].

[1] This document is a revised and extended version of [MORE96a] and [MORE96b].

1 INTRODUCTION

The aim of this work[2] is to model the process of *rational inquiry*, i.e. the (rationally controlled) transformation of the beliefs of an intelligent agent over time. The classical philosophical tradition has considered ([RESC79]) two components in this process: a *rational* one, that consists in the application of some inference procedures to the actual beliefs (resulting in the addition of inferred beliefs or the discovery of some incompatibility in them), and an *empirical* one, which adds or removes beliefs according to the result of observations performed in the agent's environment. We will also consider these two dimensions of analysis, as shown in section 3.

Logics of belief (doxastic logics, see e.g. [HAMO92]) are tools used to analyse in a formal way the reasoning about belief performed by an intelligent agent. These logics are modal logics (see [HUCR68], [JANS90]) in which the necessity operator is interpreted as belief. The formulas of these languages are usually given a semantic interpretation in the *possible worlds model* ([HINT62]), with its associated *Kripke semantics* ([KRIP63]). This semantics postulates that there is a set of *possible worlds* W and a *binary accessibility relation* R between them that connects each world w with all the worlds that could be the real one according to the agent's beliefs in w. The agent is said to believe a proposition p in a world w iff p is true in all the worlds R-accessible from w (i.e. those worlds w' s.t. (w R w')).

Some drawbacks become apparent when these logics are used to model an agent's process of rational inquiry. The main problem is that modelled agents are *logically omniscient* (they believe all classical tautologies) and *perfect reasoners* (their beliefs are closed under classical deductive closure). These properties could be appropriate if the agent's designer wished to model ideal agents (e.g. the children in the muddy children puzzle, see [FHMV95]), but they are clearly inadequate if more realistic (human or computational) agents are considered. Some arguments against logical omniscience and perfect reasoning could be the following:

- The agent can be unaware of certain facts (Fagin and Halpern tried to take this fact into account in their *logic of general awareness*, [FAHA85]).
- The agent can have limited resources (e.g. the time required or the space needed to perform a given inference). This is the most evident justification, if rational agents have to be implemented at all in real computers.
- The agent can ignore some relevant rules (e.g. the agent may not have been told what the rule of Modus Tollens is). This view was clearly considered by Konolige in his Ph.D. thesis ([KONO86]), where each agent was modelled with a base set of beliefs and a (possibly incomplete) set of inference rules.

[2]Funded by the European Union, in the Human Capital and Mobility project *VIM* (A Virtual Multi-computer), under contract ERBCHRXCT930401.

Our goal is to overcome these difficulties, by developing a way to model the process of rational inquiry (the evolution of the beliefs of a rational agent over time as a consequence of its interaction with the world and its internal inferential processes), keeping the possible worlds model and the Kripke semantics (because, after all, they seem a natural and intuitive semantics for modal logics) but trying to avoid logical omniscience and perfect reasoning.

2 SYNTACTIC VIEW OF POSSIBLE WORLDS

We define a *world* as a set of formulas in a (subset of a) propositional doxastic language. There are no conditions imposed on this set, so it can be both *partial* (most facts about the actual world will probably not be contained in each possible world) and *inconsistent* (although perhaps the inconsistency is not apparent in the set, it may be hidden in its deductive closure). It can be argued in favour of this kind of worlds with a number of ideas:

- A logically inconsistent possible world could model for instance the situation that arises when the agent is unable to take all its beliefs into account in every inference; if it focuses in a subset of them (call that a *context*), it can draw conclusions which are consistent within the context but inconsistent if all the beliefs are considered (as remarked by Shoham in [SHOH91]; Fagin and Halpern modelled this situation in their *logic of local reasoning*, [FAHA85]; Delgrande also considered this idea in [DELG95]).
- It is indeed possible to define interesting procedures of inquiry over inconsistent sets of beliefs (as shown e.g. in [RESC79]).
- The situation where no tautological information is believed could model ([JASP94]) for instance the initial state of information in the environmentalist *tabula rasa* theories of Locke and Rousseau.

Intuitively, *a world is envisioned as any situation conceivable by the agent*. This characterisation of possible worlds implies that a world can fail to contain some (even all) tautologies (and thus, logical omniscience is -in principle- avoided), and that it does not have to be deductively closed (and, therefore, the agent is not -in principle- a perfect reasoner).

In the AI literature most authors deal with these problems by introducing *impossible* possible worlds, e.g. possible worlds where logical connectives do not behave in the usual way, or tautologies may not be true, or inconsistent formulas may hold (see e.g. *impossible possible worlds* in [HINT75], *non-classical worlds* in [CRES72], *non-standard worlds* in [RESC79] or *situations* in [LEVE84]). These semantic approaches somehow alleviate the problems of logical omniscience and perfect reasoning, but they cause different problems (see e.g. [McAR88], [REIC89] for detailed reviews of these and other approaches). On the other hand, the main inconvenient of syntactic approaches (i.e. defining worlds as sets of formulas) is that they are very limiting, in the sense that the agent does not have a minimal set of inference capabilities (it can believe p and $(p \Rightarrow q)$ and not believe q, or it can believe $(p \wedge q)$ and not believe either p or q). Some authors (see e.g. [HALP86]) argue that, in this case, there seems to be no way to make a knowledge-based analysis of the agent's beliefs. The rest of this document shows how syntactically-generated beliefs can be analysed and how they evolve in time as a consequence of this analysis.

3 DYNAMIC BELIEF ANALYSIS

The evolution of the agent's beliefs over time is due to a dynamic analysis composed by two interwoven strands:

- a *logical* dimension that models the inferential (strictly deductive) processes carried out by the agent, and
- an *experimental* dimension that models the changes of beliefs caused by the interaction of the agent with its environment.

The rest of the document describes only the logical analysis (a description of the experimental side of the analysis and a whole picture of the belief system can be found in [MORE95a], [MORE95b]). The basic idea is that the experimental side of the analysis is used by the agent in order to corroborate or falsify the results obtained in the logical analysis, in a Popperian fashion ([POPP34]). If a belief obtained in the logical analysis is falsified in the experimental analysis, a belief revision procedure would be used to change the initial beliefs accordingly and the agent could proceed with the logical analysis using the revised set of beliefs.

4 LOGICAL ANALYSIS

$\mathcal{L}$ is a language composed by a finite fixed set of basic propositions ($\mathcal{P}$), two logical operators $(\neg, \vee)$[3] and the modal belief operator B. The formulas of this language are restricted to the standard propositional ones prefixed by a (possibly negated and possibly empty) sequence of modal operators[4] ($formula \Longrightarrow [\neg][B^*]propositional\_formula$). A *world* is defined as a set of formulas of $\mathcal{L}$.

In the logical analysis the agent generates a sequence of accessibility relations (see [MORE95a], [MORE95b], [MORE95c]); each of them (*via* the standard Kripke semantics) defines a different set of beliefs. This is the way in which the logical analysis models the evolution of the agent's beliefs over time due to its internal deductive inferences. This evolution will be illustrated with a small example, where the agent's beliefs in world w_e

[3] Conjunction and material implication can be defined from negation and disjunction in the usual way.

[4] Only one agent and a subset of propositional modal logic are considered in this document; both restrictions will be overcome in future work.

will be analysed. World w_e, as defined above, is just a set of formulas[5] of $\mathcal{L}$, e.g. $w_e \equiv \{\texttt{P}, \neg\texttt{Q}, (\texttt{R} \vee \texttt{S}), B\texttt{R}, B(\neg\texttt{R} \vee (\texttt{S} \vee \texttt{P}))\}$.

Using the initial set of beliefs the agent can generate the initial accessibility relation, R_0, by applying the standard Kripke semantics *backwards*: world w' is R_0-accessible from world w_e iff every formula which is believed by the agent in w_e appears in w'. Moreover, the beliefs of the agent in a world w are restricted in the following way: they have to be exactly all the propositions which appear in all its doxastic alternatives, i.e. there can be no proposition common to all the accessible worlds from w that is not a belief in w. The ontology of possible worlds just defined is represented in figure 1.

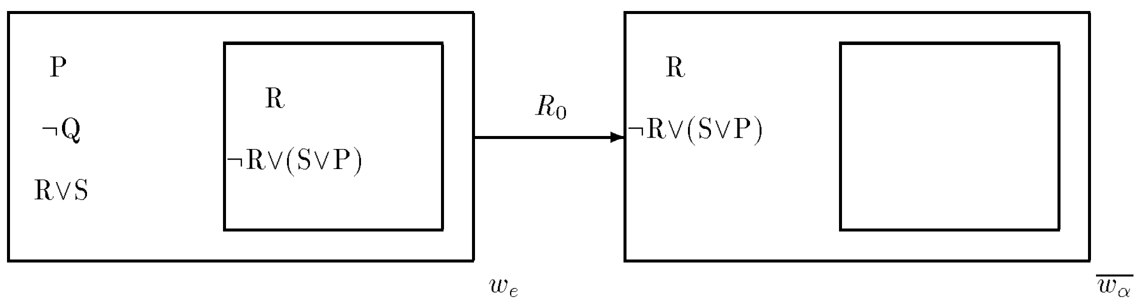

Figure 1: World w_e and its doxastic alternatives

Each world w is represented by a rectangle that contains four columns; from left to right, they represent

- non-modal formulas that must be contained in w,
- non-modal formulas that must *not* be contained in w,
- formulas that are believed in w and
- formulas that are *not* believed in w[6].

Classes of worlds ($\overline{w}$) are represented in the same way (reading *contained/believed in w* as *contained/believed in all the worlds in class $\overline{w}$*). w_e is R_0-related to all the worlds in class $\overline{w_\alpha}$, i.e. all the worlds that contain (at least) R and $(\neg\texttt{R} \vee (\texttt{S} \vee \texttt{P}))$.

The logical analysis of the agent's beliefs is performed using (a modified version of) the *analytic tableaux method* (see e.g. [SMUL68]). A tableau is divided into a left and a right column[7]. The analysis starts with a tableau T_0 that contains in its left column the initial

[5] Nested beliefs are not considered in this example; they would be analysed in a similar way (an analytic tableaux tree would be necessary for each level of nested beliefs). The example's initial set of beliefs does not contain negated beliefs either; their treatment appears to be tricky.

[6] The last two columns are included in an inner rectangle to reinforce somehow its interpretation as beliefs. In figure 1 the second and the fourth columns in w_e are empty, and only the first column in $\overline{w_\alpha}$ is used.

[7] This is the first difference with respect to classical analytic tableaux, that only contain one set of formulas.

set of beliefs of the agent in w_e: {R, (¬R ∨ (S ∨ P))}. In the light of previous considerations, it can be argued that T_0 not only represents the agent's beliefs in w_e; it can also be seen as a (partial) description of all the worlds initially connected to w_e, because all the beliefs in w_e are true in all of these worlds (T_0 represents $\overline{w_\alpha}$). Thus, a tableau represents a class of worlds: all the worlds in which the propositions contained in the tableau are true. More precisely, *a tableau represents all those worlds that contain the formulas of its left column and do not contain any of the formulas of its right column.*

The use of explicit negative information in the analytic tableaux somehow enriches the kind of situations that can be modelled with the standard possible worlds model and its Kripkean semantics. Given a formula α, in the classical picture (where negative information is implicit, as everything that is not included in the tableaux), only three situations are possible: $\{(B\alpha, \neg B\neg\alpha), (B\neg\alpha, \neg B\alpha), (\neg B\alpha, \neg B\neg\alpha)\}$; if explicit (positive and negative) information is considered in every tableau, you can model those situations plus the following ones:

- (): the agent has never thought about α.
- $(B\alpha)$, $(B\neg\alpha)$: these situations could appear as intermediate steps towards the more complete stages $(B\alpha, \neg B\neg\alpha)$ and $(B\neg\alpha, \neg B\alpha)$.
- $(\neg B\alpha)$, $(\neg B\neg\alpha)$: in these situations the agent does not believe something, but it has not (yet) thought whether it believes its negation or not[8].

The *splitting rule* of the analytic tableaux method has also been modified[9]: when it is applied to analyse a disjunction $(\phi \vee \psi)$ contained in the left column of a tableaux T, it *generates three subtableaux*: the first one has ϕ and ψ in its left column; the second one has ϕ in its left column and ψ in its right column, and the third one has ψ in its left column and ϕ in its right column. Thus, with this rule the agent explores all the possibilities of accounting for the truth of the analysed disjunction, namely that one (or both) of its members are true. The three subtableaux also keep all the formulas contained in T[10].

After the application of this rule to analyse (¬R ∨ (S ∨ P)) (the only disjunction in T_0), the situation shown in figure 2 is reached.

Consider the following partition of class $\overline{w_\alpha}$:

- $\overline{w_{\alpha_1}}$: worlds of $\overline{w_\alpha}$ that contain ¬R and (S ∨ P).
- $\overline{w_{\alpha_2}}$: worlds of $\overline{w_\alpha}$ that contain ¬R but do not contain (S ∨ P).

[8]Note that, in our framework, $\neg B\alpha$ does *not* imply $B\neg\alpha$.

[9]Recall that, in the classical analytic tableaux method, the analysis of a disjunction $(\alpha \vee \beta)$ causes the generation of two subtableaux, one containing α and the other one containing β.

[10]This is a minor modification with respect to the classical method; it is nevertheless convenient because later we can talk about *formulas contained in all open tableaux* instead of the longer and more cumbersome *formulas that are contained in every branch of the tableaux tree.*

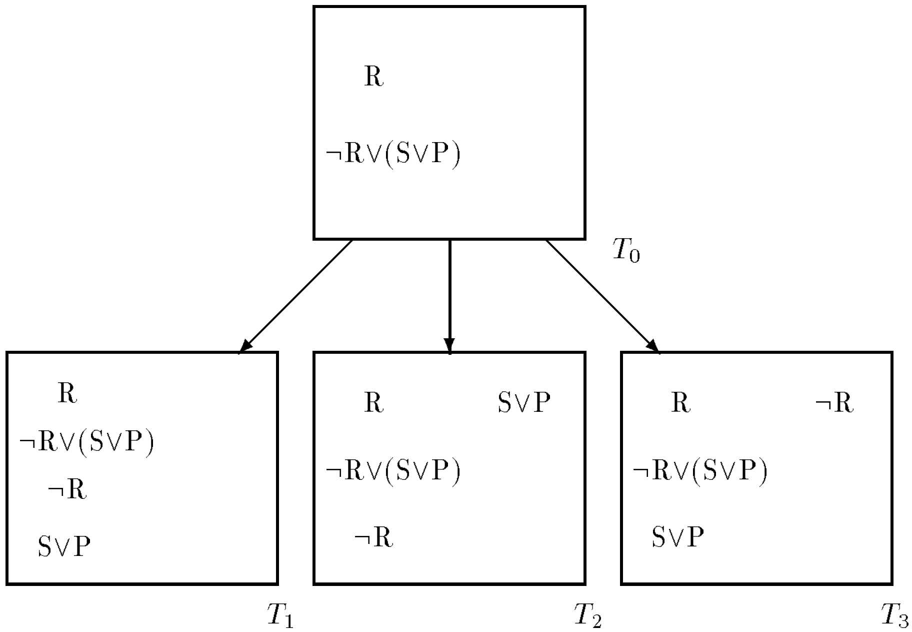

Figure 2: Splitting rule applied to $(\neg \mathtt{R} \vee (\mathtt{S} \vee \mathtt{P}))$

- $\overline{w_{\alpha_3}}$: worlds of $\overline{w_\alpha}$ that contain $(\mathtt{S} \vee \mathtt{P})$ but do not contain $\neg\mathtt{R}$.
- $\overline{w_{\alpha_4}}$: worlds of $\overline{w_\alpha}$ that do not contain either $\neg\mathtt{R}$ or $(\mathtt{S} \vee \mathtt{P})$.

The tableau T_1 represents the class of worlds $\overline{w_{\alpha_1}}$, T_2 represents $\overline{w_{\alpha_2}}$ and T_3 represents $\overline{w_{\alpha_3}}$[11]. The worlds in class $\overline{w_{\alpha_4}}$ are not even considered by the analytic tableaux method, because it looks for models of the initial set of formulas and it is not possible to have a model of the set $\{\mathtt{R}, (\neg\mathtt{R} \vee (\mathtt{S} \vee \mathtt{P}))\}$ in which both $\neg\mathtt{R}$ and $(\mathtt{S} \vee \mathtt{P})$ fail to be true.

In the tableaux analysis that is being considered, a tableau is *closed* if any of the following circumstances holds:

- The tableau contains a formula and its negation[12] in its left column; the worlds belonging to the class of worlds represented by the tableau contain an explicit contradiction, so they are considered *logically impossible* and eliminated from the analysis.
- The tableau contains the same formula both in its left and right columns[13]; it represents the worlds that contain and do not contain that formula, so it represents the empty set of worlds and is dismissed from the analysis.

[11]This is another reason to consider tableaux with positive and negative information; it is easier to picture the analysis of a disjunction as a partition in the class of doxastic alternatives; in the classical method, the classes of worlds represented by the two subtableaux generated by the splitting rule overlap, i.e. there are worlds that belong to both classes (the worlds that contain both disjuncts).

[12]This is also the closing condition in the classical method.

[13]This situation does not arise in the example used in this document; it would for instance if R were replaced by $\neg R$.

In the example, T_1 and T_2 contain both R and ¬R in their left columns, so the agent would consider the classes of worlds represented by these tableaux ($\overline{w_{\alpha_1}}$, $\overline{w_{\alpha_2}}$) as logically impossible and it would get rid of these worlds by closing T_1 and T_2 and dismissing the branches that contain them from the logical analysis. The remaining tableaux tree contains only T_0 and T_3. Thus, the effect of the analysis of the formula (¬R ∨ (S ∨ P)) is the restriction of the doxastic alternatives of w_e, that shrink from $\overline{w_\alpha}$ to $\overline{w_{\alpha_3}}$. The application of the splitting rule is modelling the generation of a new accessibility relation, R_1, that restricts the set of R_0-accessible worlds. This restriction is shown in figure 3.

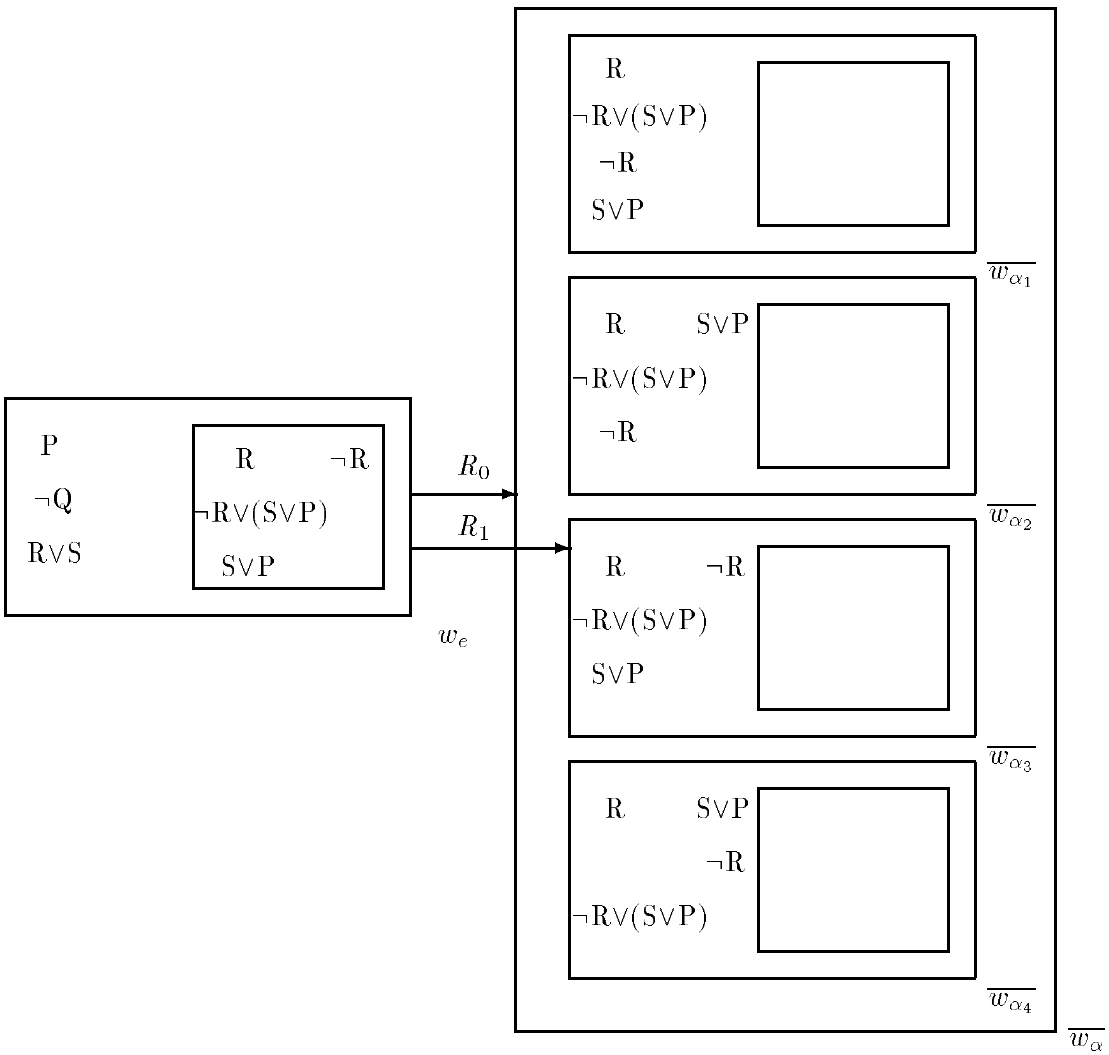

Figure 3: Reduction of belief alternatives

w_e was R_0-related to all the worlds in class $\overline{w_\alpha}$, but is R_1-related only to the worlds in class $\overline{w_{\alpha_3}}$. The change of the accessibility relation implies a possible change in the beliefs of the agent. The beliefs in w_e at this point would be {R, (¬R ∨ (S ∨ P)), (S ∨ P)}. Notice that the restriction of the belief alternatives has caused the addition of a new belief, (S $\vee$ P). This formula is a new belief because it was not included in all the worlds in $\overline{w_\alpha}$ but it appears in all the worlds in $\overline{w_{\alpha_3}}$. Furthermore, as $\neg R$ is *not* included in any of the worlds in $\overline{w_{\alpha_3}}$, the analysis also shows that it is *not* a belief of the agent in w_e, and this fact is also reflected in figure 3, with the addition of $\neg R$ in the fourth column of the representation of w_e[14].

If the agent wants to pursue the analysis of its beliefs it can generate the next accessibility relation, R_2, by analysing the disjunction (S $\vee$ P) in T_3, obtaining the result shown in figure 4.

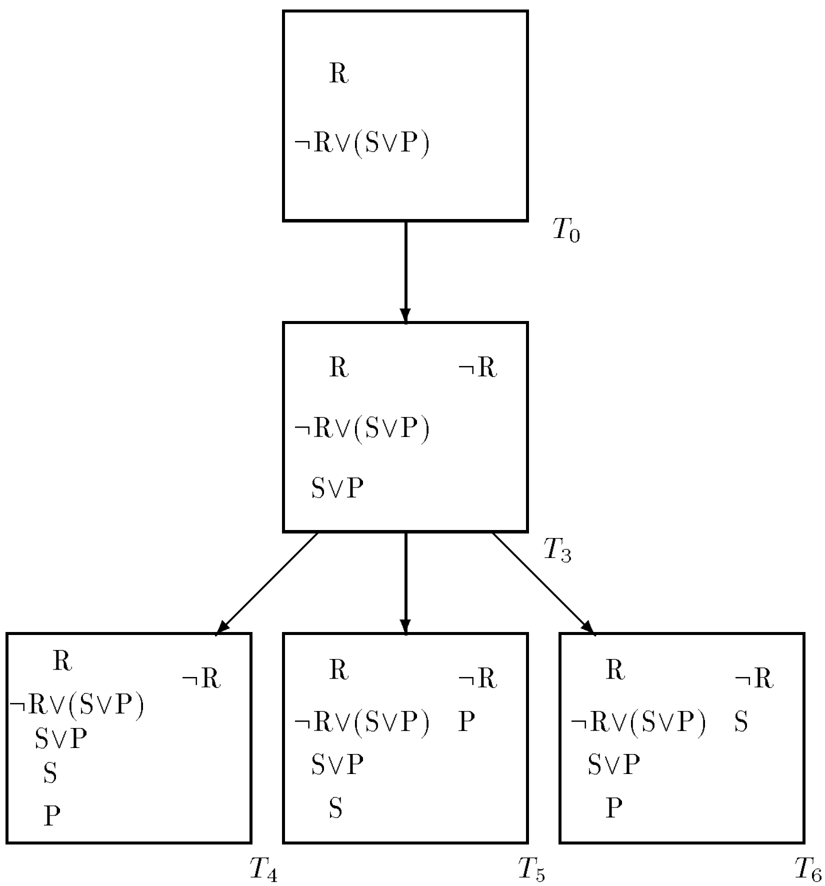

Figure 4: Analysis of (S $\vee$ P)

A more intuitive interpretation of the restriction from R_1 to R_2 results if the following partition of $\overline{w_{\alpha_3}}$ in four subclasses is considered:

- $\overline{w_{\alpha_{3_1}}}$: worlds of $\overline{w_{\alpha_3}}$ that contain S and P.

[14]Summarising, positive beliefs are those formulas contained in the first column of all classes of doxastic alternatives, whereas negative beliefs are those formulas contained in the second column of at least one class of doxastic alternatives.

- $\overline{w_{\alpha_{3_2}}}$: worlds of $\overline{w_{\alpha_3}}$ that contain `S` but do not contain `P`.
- $\overline{w_{\alpha_{3_3}}}$: worlds of $\overline{w_{\alpha_3}}$ that contain `P` but do not contain `S`.
- $\overline{w_{\alpha_{3_4}}}$: worlds of $\overline{w_{\alpha_3}}$ that do not contain either `S` or `P`.

w_e, that was R_1-related to the worlds in $\overline{w_{\alpha_3}}$, is only R_2-related to the worlds of the subclasses $\overline{w_{\alpha_{3_1}}}$, $\overline{w_{\alpha_{3_2}}}$ and $\overline{w_{\alpha_{3_3}}}$, but not to those in $\overline{w_{\alpha_{3_4}}}$ (classes $\overline{w_{\alpha_{3_1}}}$, $\overline{w_{\alpha_{3_2}}}$ and $\overline{w_{\alpha_{3_3}}}$ are represented by tableaux T_4, T_5 and T_6). This situation is depicted in figure 5.

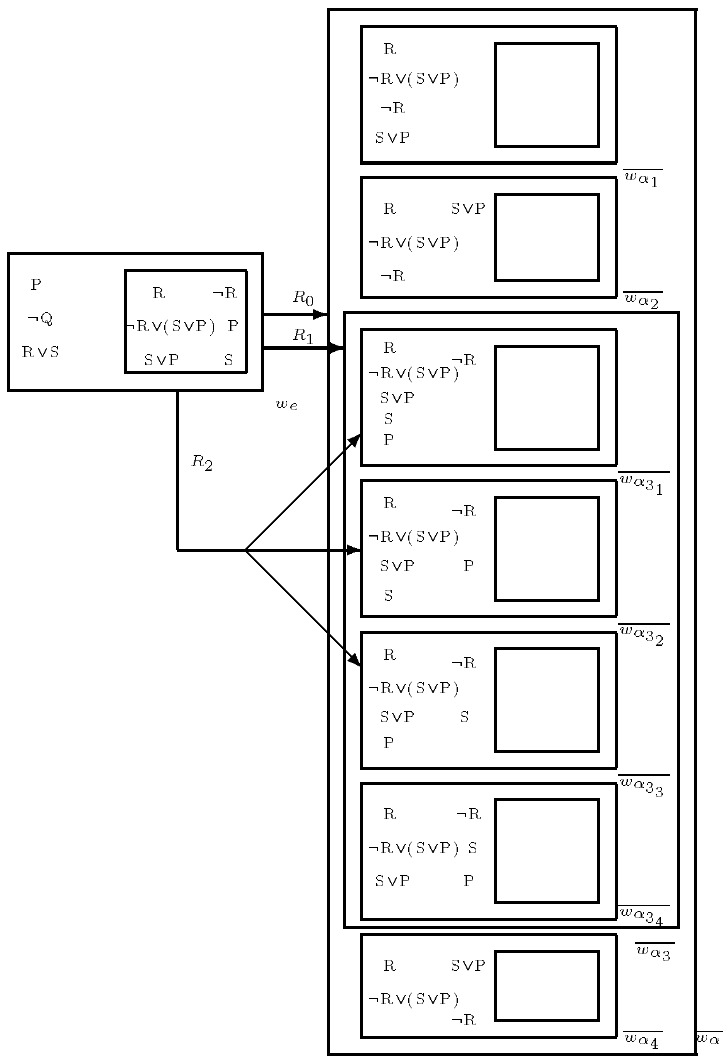

Figure 5: Further reduction of belief alternatives

Notice that the set of positive beliefs has not changed, because all the formulas common to all the R_2-accessible worlds were already common to all the R_1-accessible worlds. However, the set of negative beliefs has increased, because `P` is not contained in all the doxastic alternatives (because the worlds in $\overline{w_{\alpha_{3_2}}}$ do not contain it). The same situation happens to `S`, which is not contained in the worlds in class $\overline{w_{\alpha_{3_3}}}$; thus, neither `P` nor `S` are believed in w_e. This fact is reflected in the fourth column of the rectangle that represents w_e in figure 5. There are no more disjunctions to be analysed in the open tableaux, so this is the end of the (purely logical) analysis that the agent can perform in w_e (without introducing instances of tautologies, topic considered in section 5).

5 INTRODUCING TAUTOLOGIES

As shown in section 4, the classical analytic tableaux method has been modified in a number of ways:

- A tableau contains two sets of formulas (the left and the right columns). It represents the set of possible worlds that contain all the formulas in its left column and do not contain any of the formulas in its right column.
- When the splitting rule[15] is applied, three subtableaux are generated. One of them contains both members of the analysed disjunction in its left column, while the other two contain one of them in the left column and the other one in the right column. These subtableaux represent all those (classes of) worlds in which the analysed disjunction holds.
- After applying a splitting rule, the analysed proposition and all the other formulas in the tableau are maintained in the resulting tableaux, they are not deleted. The agent would just add a tag to the proposition in other to recall that it had already been analysed.

There is yet another important modification to the classical analytic tableaux method to be mentioned: *the agent is* **not** *allowed to add to an open tableau any tautology* (recall that, in the classical analytic tableaux method, a tautology can be introduced in any open tableau at any point). Notice that the agent's (positive) beliefs at a certain stage of the analysis are those formulas that appear in all open tableaux. Thus, if the agent could freely introduce any tautology in any open tableau, logical omniscience could not be avoided. Moreover, in that case the agent would also be able to reach any deductive consequence of its initial set of beliefs, so it would also be a perfect reasoner[16].

[15]The splitting rule previously described is used to analyse disjunctions in the left column; five more rules are needed in order to analyse all the formulas that can be generated using negation and disjunction; i.e. we need to be able to analyse formulas of the form $(\alpha \vee \beta)$, $\neg\neg\alpha$ and $\neg(\alpha \vee \beta)$, in both columns. In the rest of this document only clauses (and initial positive beliefs) will be considered; therefore, only the disjunction splitting rule -with a brief exception to be explained later- will be used.

[16]If a formula α is a logical consequence of the initial set of beliefs I, it may be obtained in the logical analysis as a new belief by introducing the (tautological) formula $(\alpha \vee \neg\alpha)$ in all open tableaux and exhaustively analysing all those tableaux containing I and $\neg\alpha$, until all of them are closed. That situation would eventually be reached because the set $I \cup \{\neg\alpha\}$ is inconsistent.

Forbidding the use of tautologies is, however, a very drastic way of avoiding logical omniscience and perfect reasoning, because it puts severe limits to the agent's deductive capabilities. Consider the following definition:

Definition 1 *A formula f is derivable from an initial set of clauses I[17] ($I \models_0 f$) iff it is possible to perform a logical analysis, using the analytic tableaux method as described in section 4, with the initial tableau T_0 containing I in its left column, in such a way that f appears in the left columns of all open tableaux (at least one[18]).*

Recall that the open tableaux represent classes of doxastic alternatives; therefore, if at a certain point of the logical analysis all open tableaux contain a certain formula f, this formula is contained in all the doxastic alternatives so, *via* the Kripke semantics, it is a belief of the agent at that point. Therefore, the set of potential beliefs is equal to the set of derivable formulas. This set can be characterized ([MORE95b]) in the following way:

Definition 2 *Let Δ be a set of clauses and let c_1, c_2 be clauses.*

- *SUBSUMED(Δ, c_1) = $\{c_2 \epsilon \Delta$ s.t. c_2 is subsumed by $c_1\}$*
- *REST(Δ, c_1) = $\{c_2$ s.t. $(c_1 \vee c_2)$ ϵ SUBSUMED(Δ, c_1)$\}$*
- *DELETE(Δ, c_1) = Δ - SUBSUMED(Δ, c_1) $\cup$ REST(Δ, c_1)*

Lemma 1 *($I \models_0 f$) iff*

- *I does not contain a formula and its negation, and*
- *I does not contain the negation of f, and*
- - *($f \epsilon I$) or*
  - *SUBSUMED(I, f)$\neq\emptyset$ and DELETE(I, f) is inconsistent.*

The first and the second conditions are trivial (they will be called TC in the rest of the document): if I contained a formula and its negation, T_0 would automatically be closed, so the set of derivable beliefs from I would be empty; if I contained a given formula, p, then $\neg p$ could not be derived, because as soon as it were generated, it would create a tableau containing p (that would be contained in all tableaux) and $\neg p$, and therefore it would immediately be closed. The third condition states that f must be a clause of I or a subclause such that, if it is removed from every clause in which it appears, it generates an inconsistent set. Assume that $I = \Delta \cup \{(f \vee c_1), ..., (f \vee c_p)\}$, where the formulas in

[17]For the sake of simplicity, it will be assumed in the rest of the document that all the formulas in I are clauses (I=$\{c_1, c_2, \ldots, c_n\}$); i.e. the initial set of formulas has been transformed into conjunctive normal form.

[18]There must be at least one open tableau, otherwise *any* formula would be derivable from an inconsistent set, which is a consequence that we want to avoid (i.e. we want our model to allow some paraconsistency).

Δ do not contain f. f can be derived from I iff the tableau containing $\Delta \cup \{c_1, ..., c_p\}$ is closed, and that will happen iff that set is inconsistent.

Therefore, if no tautologies are allowed, the set of derivable formulas is quite small (a subset of the subclauses that are logical consequences[19] of the initial set of clauses); on the other hand, if any tautology is allowed, logical omniscience and perfect reasoning are unavoidable. This fact is suggesting an interesting way of controlling the agent's deductive capabilities: *to allow the use of some (instances of) tautologies in the logical analysis*. For instance ([MORE95b],[MORE95c]), the agent could use some instances of the *Axiom of the Excluded Middle*, AEM, $(p \vee \neg p)$. The use of this particular tautology seems a natural way to allow the agent to have *doubts*, to wonder whether it believes some formula (p) or its negation $(\neg p)$. This exception permits the introduction of the formula $(p \vee \neg p)$ in a tableau, which is later split into two subtableaux containing p in one column and $\neg p$ in the other[20]. In this way, the agent can explore both alternatives independently, and the logical analysis can guide the search of examples or counter-examples needed to give more credence to one side of the doubt than to the other. In fact, the possibility of adding this kind of tautologies in the analytic tableaux is a well-known idea in the tradition of classical proof theory, as described *e.g.* in [BELL77]. It has also been suggested by Hintikka in his *interrogative model of inquiry* (see e.g. [HINT81], [HINT86], [HINT92]), in which he models the process of scientific inquiry with plays of the *interrogative game*, which is a game played by the *Inquirer* and *Nature*. The Inquirer seeks to know whether a certain conclusion C is true or not, and it can perform *deductive* moves (controlled with analytic tableaux) and *interrogative* moves, which are questions put to Nature. In these interrogative moves, Hintikka allows the introduction of (some) instances of AEM in the left side[21] of the subtableaux (those formulas $(p \vee \neg p)$ in which p is a subsentence of a formula of the subtableaux or a substitution instance of a subformula of a formula in the subtableaux with respect to names appearing in the subtableaux).

The introduction of instances of AEM is also *a way to control the set of formulas that can be believed by the agent*. The set of derivable formulas depends on the kind of instances of AEM that are allowed, so putting restrictions on the use of this principle is a way to constrain the set of obtainable formulas (the set of potential beliefs). For instance, assume a less restrictive derivability relation, $\models_s$, defined as follows:

Definition 3 *Subclause rule: the agent is allowed to introduce at any open tableau in the logical analysis those instances of AEM* $(s \vee \neg s)$ *in which* s *is a subclause of a clause of the initial set of beliefs.*

Subclause-extended logical analysis: logical analysis in which the agent is allowed to use the subclause rule.

[19] If f is derivable, the tableau $I \cup \{\neg f\}$ would close in the classical analytic tableaux method, simply by analysing the same clauses that had been analysed to proof the derivability of f.

[20] The third subtableau would be immediately closed because it would contain both p and $\neg p$ in its left column.

[21] Hintikka uses the classical Beth-style analytic tableaux, as described in [BETH55].

A formula f is s-derivable from an initial set of clauses I ($I \models_s f$) iff it is possible to perform a subclause-extended logical analysis in such a way that f appears in the left columns of all open tableaux (at least one).

The set of s-derivable formulas is larger than the set of derivable formulas, as shown in the following characterisation[22] of the former set:

Lemma 2 *($I \models_s f$) iff TC and*

- *$I \models f$, being f a subclause of a clause of I, or*
- *f=($s \vee \neg s$), being s a subclause of a clause of I*

In other words, the agent can s-derive (excluding the trivial conditions TC and the last clause of the previous lemma, that is also obvious) a subclause s of I iff $I \models s$. It could s-derive that formula by introducing ($s \vee \neg s$) in all open tableaux and analysing (and eventually closing) all those tableaux that do not contain the formula s in its left column. For instance, the formula ($A \vee B$) is s-derivable (but not derivable) from the set $\{(D \vee A), (\neg D \vee B), ((A \vee B) \vee C)\}$.

This kind of ideas can be pushed even further, when the use of different tautologies is considered (see [MORE95b] for a preliminary study of the use of different instances of AEM and different tautologies in the logical analysis). In fact, the main point of the logical analysis is that not only logical omniscience and perfect reasoning are avoided, but also *the potential beliefs of the agent can be controlled by restricting the (instances of the) tautologies that the agent can use in the logical analysis*.

The logical analysis suggested in this document also introduces another way of modelling an imperfect reasoner (even if it is allowed to use any tautology). The classical analytic tableaux method is refutative, in the sense that, in order to prove that α is a consequence of a set I, it tries to fail in building a model of $I \cup \{\neg\alpha\}$, by *exhaustively*[23] analyzing the tableau containing I and $\neg\alpha$ and closing all the subtableaux in the analytic tableaux tree. Our analysis has an important difference: the agent is *not* obliged to exhaustively analyze any tableau. In this way we can model imperfect reasoning due to a lack of resources, whereas with the previous kind of imperfection we were modelling those agents that do not have a perfect deductive power. The agent may decide to stop the logical analysis at any point of the derivation tree, and to proceed with other tasks (e.g. to make some experimental tests in order to corroborate or falsify the current beliefs obtained in the logical analysis).

---

[22]In these derivations the agent will need to use once an extra rule: a tableau containing the formula $\neg(\alpha \vee \beta)$ in its left column can generate a subtableau containing the sames formulas plus $\neg\alpha$ and $\neg\beta$.

[23]The tableaux obtained in the course of the analysis are just Hintikka's model sets waiting to be completed; they are not models (yet), so they are not useful to the agent.

6 SUMMARY AND FUTURE WORK

In this article a two-dimensional dynamic analysis of a rational agent's set of beliefs has been suggested, and its logical dimension has been summarily described. The logical analysis is performed using a modified version of the analytic tableaux method, and it allows the agent to deduce some facts from its initial set of beliefs. Logical omniscience and perfect reasoning are avoided in two ways:

- The deductive power of the agent can be controlled through appropriate restrictions on the instances of tautologies that it is allowed to introduce in the open tableaux.
- The agent may decide not to perform an exhaustive logical analysis, and to consider intermediate (positive and negative) beliefs as good enough to guide its behaviour or to serve as the basis of experimental analysis.

In the future work these constraints will be studied to discover which (instances of) tautologies yield interesting sets of deductive consequences. Another key issue in this research will be the study of the interaction between the two strands of analysis (see [MORE95b] for a preliminary study of these topics).

References

[BELL77] Bell, J., Machover, M., *A Course in Mathematical Logic*, North Holland, 1977.

[BETH55] Beth, E., 'Semantic entailment and formal derivability', in *Mededelingen van de Koninklijke Nederlandse Akademie van Wetenschappen, Afdeling Letterkunde* 13, 309-342, 1955.

[CRES72] Cresswell, M., 'Intensional logics and logical truth', *Journal of Philosophical Logic* 1, 2-15, 1972.

[DELG95] Delgrande, J., 'A framework for logics of explicit belief', *Computational Intelligence* 11, 47-86, 1995.

[FAHA85] Fagin, R., Halpern, J., 'Belief, awareness and limited reasoning', *Proceedings of IJCAI-85*, 491-501, 1985.

[FHMV95] Fagin, R., Halpern, J., Moses, Y., Vardi, M., *Reasoning about Knowledge*, MIT Press, 1995.

[HALP86] Halpern, J., 'Reasoning about knowledge: an overview', *Proceedings of TARK-86*, 1-17, 1986.

[HAMO92] Halpern, J., Moses Y., 'A guide to completeness and complexity for modal logics of knowledge and belief', *Artificial Intelligence* 54, 319-379, 1992.

[HINT62] Hintikka, J., *Knowledge and Belief*, Cornell University Press, Ithaca, N.Y., 1962.

[HINT75] Hintikka, J., 'Impossible possible worlds vindicated', *Journal of Philosophical Logic* 4, 475-484, 1975.

[HINT81] Hintikka, J., 'On the logic of an interrogative model of scientific inquiry', *Synthese* 47, 69-83, 1981.

[HINT86] Hintikka, J., 'Reasoning about knowledge in philosophy: the paradigm of epistemic logic', *Proceedings of TARK-86*, 63-80, 1986.

[HINT92] Hintikka, J., 'The interrogative model of inquiry as a general theory of argumentation', *Communication and Cognition* 25, 221-242, 1992.

[HUCR68] Hughes, G., Cresswell, M., *An Introduction to Modal Logic*, Methuen and Co., 1968.

[JANS90] Jansana, R., *Una introducción a la lógica modal*, Ed. Tecnos, 1990.

[JASP94] Jaspars, J., *Calculi for Constructive Communication*, Institute for Logic, Language and Computation, ILLC Dissertation Series 1994-4, 1994.

[KONO86] Konolige, K., *A Deduction Model of Belief*, Morgan Kaufmann, 1986.

[KRIP63] Kripke, S., 'A semantical analysis of modal logic I: normal modal propositional calculi', *Zeitschrift für Mathematische Logik und Grundlagen Mathematik* 9, 67-96, 1963.

[LEVE84] Levesque, H., 'A logic of implicit and explicit belief', *Proceedings of AAAI-84*, 198-202, 1984.

[McAR88] McArthur, G., 'Reasoning about knowledge and belief: a survey', *Computational Intelligence* 4, 223-243, 1988.

[MORE95a] Moreno, A., Sales, T., 'Dynamic belief modeling', *Fourth International Colloquium on Cognitive Science*, Donostia, May 1995. An extended version is available as Research Report 95-28-R, Department of Software, Technical University of Catalonia (UPC), 1995 (http://www-lsi.upc.es/www/dept/techreps/techreps.html).

[MORE95b] Moreno, A.,'Dynamic belief analysis', Research Report 73, Human Communication Research Centre, University of Edinburgh, 1995 (http://www.cogsci.ed.ac.uk/hcrc/publications/).

[MORE95c] Moreno, A.,'Análisis dinámico de las creencias', *Proceedings of the VI Conference of the Spanish Association for AI, CAEPIA-95*, 103-113, (1995).

[MORE96a] Moreno, A.,'Dynamic belief analysis', submitted to *ECAI-96*.

[MORE96b] Moreno, A.,'Limited logical belief analysis', submitted to *IBERAMIA-96*.

[POPP34] Popper, K., *The Logic of Scientific Discovery*, Hutchinson and Co. Ltd., 1934.

[REIC89] Reichgelt, H., 'Logics for reasoning about knowledge and belief', *Knowledge Engineering Review* 4, 119-139, 1989.

[RESC79] Rescher, N., Brandom, R., *The Logic of Inconsistency*, Rowman and Littlefield Eds., 1979.

[SHOH91] Shoham, Y., 'Varieties of context', in *Artificial Intelligence and Mathematical Theory of Computation*, 393-408, Academic Press, 1991.

[SMUL68] Smullyan, R., *First-order logic*, Springer Verlag, 1968.