Modelling rational inquiry in non-ideal agents

Antonio Moreno
Escola Tècnica Superior d'Enginyeria
Departament d'Enginyeria Informàtica
Universitat Rovira i Virgili (URV)
Carretera de Salou, s/n. 43006-Tarragona
amoreno@etse.urv.es

Abstract

The construction of *rational agents* is one of the goals that has been pursued in Artificial Intelligence (AI). In most of the architectures that have been proposed for this kind of agents, its behaviour is guided by its *set of beliefs*. In our work, *rational agents* are those systems that are permanently engaged in the process of *rational inquiry*; thus, their beliefs keep evolving in time, as a consequence of their internal inference procedures and their interaction with the environment. Both AI researchers and philosophers are interested in having a formal model of this process, and this is the main topic in our work.

Beliefs have been formally modelled in the last decades 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*. We avoid these problems by replacing possible worlds by *conceivable situations*, which are all the situations that the modelled agent is capable of considering.

In this document we show how this notion of *conceivable situations* may be used to model the process of rational inquiry in which a non-ideal rational agent is engaged. We define a wide class of agents, called *rational inquirers*, which are a general abstraction of any kind of non-ideal agent. We show how the beliefs of this kind of agents evolve in time as a consequence of a multi-dimensional belief analysis, and we use the framework of *conceivable situations* in order to model this evolution.

Keywords: rational inquiry, doxastic logics, logical omniscience, perfect reasoning, possible worlds, Kripke semantics, dynamic multi-dimensional belief analysis

1 Aim of the work

1.1 Introduction

Artificial Intelligence (AI) has pursued the goal of constructing *rational agents* for a long time. This kind of agents may be defined ([RuNo95]) as *systems that have some kind of perception and try to act upon the environment so as to achieve their goals, given their beliefs*; thus, beliefs somehow guide their behaviour (*e.g.* they may be used to choose between alternative courses of action available to the agent). The architecture of such agents usually includes a knowledge base (that stores relevant facts about the agent and its environment) and modules that may perform inferences from those facts, interact with the agent's environment, create and evaluate different plans, *etc.* In the last decade it has been argued that, if the agent is to display a rational behaviour, it needs to have a module capable of handling its beliefs.

In this document we take the expression *"rational agents"* to refer to those agents that, apart from complying with the previous definition, are permanently engaged in a process of *rational inquiry* (defined by Brandom in [ReBr79] as *the [rationally controlled] transformation of belief over time*). Intuitively, these agents are constantly trying to make their beliefs as similar as possible to the facts that hold in the real world. They keep trying to expand their beliefs (by including facts that are true in their environment) and to get rid of wrong beliefs (those that do not reflect the state of the actual world). The classical philosophical tradition has considered ([ReBr79]) 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. An important concern both in AI and in Philosophy is how to *build a formal model of rational inquiry*; this is the aim of our work.

1.2 Formal models of belief

1.2.1 Possible worlds and Kripke semantics

The natural start point in our work was to consider *logics of knowledge and belief* (epistemic and doxastic logics, [HaMo92]). These modal logics are used to analyse in a formal way the reasoning about knowledge or belief performed by an agent. In the literature of doxastic logics the universal modal operator ($\Box$) is usually called B. If several (m) agents are taken into account, a family of subscripted operators (B_1, B_2, ..., B_m) is considered (where $B_i\varphi$ is read as "*Agent i believes* φ"). The usual language of propositional doxastic logic for m agents contains a set of primitive propositions (P, Q, R, ...), the basic logical operators ($\neg$, $\vee$, $\wedge$ and $\Rightarrow$) and the modal belief operators B_1, B_2, ..., B_m. The formulas of this language are the primitive propositions and the applications of the logical operators or the modal operators to other formulas of the language.

The semantic model traditionally adopted as a basis in doxastic logics is the *possible worlds model* ([Hint62]). This model is based on the assumption that there is a set of possible states (or *possible worlds*) in which the agent can be in any moment; when the agent is in a possible world, there is a set of possible worlds which are compatible with the actual world, in the sense that the agent cannot distinguish these worlds from the actual one. The usual semantics given to the formulas of the doxastic language described above is the *Kripke semantics* ([Krip63]), that states that the agent believes a formula iff it is true in all the worlds that the agent cannot tell apart from the actual world.

A *normal Kripke structure* is a tuple of the form (S, π, R_1, ..., R_m), where S is the set of possible worlds, π is a truth assignment to each primitive proposition in each world and R_i is the accessibility relation between worlds for agent i ($(s, t) \in R_i$ iff s and t are indistinguishable worlds for agent i). Given a normal Kripke structure M, the relation $M, s \models \varphi$ (read "*φ is true (or satisfied) in state s of model M*") is usually defined in the following way:

- $M, s \models$ P, being P a primitive proposition, if $\pi(s, \mathtt{P}) = true$
- $M, s \models \neg\varphi$ if $M, s \not\models \varphi$
- $M, s \models (\varphi \vee \psi)$ if $M, s \models \varphi$ or $M, s \models \psi$
- $M, s \models (\varphi \wedge \psi)$ if $M, s \models \varphi$ and $M, s \models \psi$
- $M, s \models (\varphi \Rightarrow \psi)$ if $M, s \models \neg\varphi$ or $M, s \models \psi$
- $M, s \models B_i\varphi$ if $M, t \models \varphi$ $\forall t$ such that $(s, t) \in R_i$

The last clause formalizes the conception of beliefs previously stated: an agent i believes a proposition φ when it is true in all the worlds considered possible by the agent, *i.e.* in all the worlds that the agent cannot tell apart from the actual world. That means that φ has to be true in all the worlds connected to the actual world through R_i, which is the accessibility relation between worlds for agent i.

There exists ([HaMo92]) a *sound and complete axiomatization* of the class of all normal Kripke structures for m agents. An axiomatic system S is *sound* with respect to a class of structures C if every formula provable in S is true in C. S is *complete* with respect to C if every formula which is true in C can be proved in S. An axiomatic system *characterizes* a class of structures when it is a sound and complete axiomatization of the class. The axiomatic system that characterizes the set of all normal Kripke structures for m agents has two axioms and two inference rules, which are the following:

- A1. All the instances of tautologies of propositional calculus.
- A2. $(B_i\varphi \wedge B_i(\varphi \Rightarrow \psi)) \Rightarrow B_i\psi$ (axiom K)
- R1. From $\vdash \varphi$ and $\vdash (\varphi \Rightarrow \psi)$ infer $\vdash \psi$ (*Modus Ponens*)
- R2. From $\vdash \varphi$ infer $\vdash B_i\varphi$ (Necessitation)

This axiomatic system is known as system K_m (or K if only one agent is considered), and it is the simplest one used to model logics of knowledge and belief.

1.2.2 Logical omniscience and perfect reasoning

Axiom A1 (all the instances of all propositional tautologies) and rule R1 (*Modus Ponens*) are taken directly from classical propositional logic. The problems to be addressed in this work derive from axiom K and the rule of necessitation: they (seem to) commit us to model agents that are

- **logically omniscient**, because they believe all tautologies (since all of them are true in every world), and
- **perfect reasoners**, because they also believe all logical consequences of their beliefs (*e.g.* if an agent believes P and (P $\Rightarrow$ Q) in a state s, it means that these two propositions are true in all the worlds compatible with s (all states R_i-accessible from s); therefore, Q will also be true in all of these worlds, and the agent will also believe Q).

These facts imply that an agent with the basic arithmetic axioms would have to know whether the Fermat theorem is indeed a theorem or not, or that an agent that knew the rules of chess would have to know whether White has a winning strategy or not ([Kono86]). The union of these problems is usually referred to in the literature as the problem of *logical omniscience* (even though some authors prefer to call it *closure under logical consequence* ([Kono86]) or *tautological closure* ([Shoh93])). We will distinguish between logical omniscience and perfect reasoning, as described above.

Obviously there are many circumstances in which these conditions are unacceptable; that would be the case when the agent is supposed to be able to compute its knowledge or to take actions based on it. This would be an *internal* view of knowledge, as something that is acquired after a computation. It is clearly not a realistic model of either human agents (who are not logically omniscient) or computational agents (which have resource limitations that can prevent them from being perfect reasoners). In summary ([Perl94]), omniscience is irreparably out of line with the needs of any real reasoning agent.

We may now state a more refined version of the aim of this work: to develop a way to *model the process of rational inquiry* (the evolution of a rational agent's set of beliefs 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 very natural and intuitive semantics for modal logics of belief) *but trying to avoid the problems of logical omniscience and perfect reasoning* (in order to take into account non-ideal agents).

2 Modelling tools

2.1 Conceivable situations

Several authors have tried to solve (or, at least, partially alleviate) the problems of logical omniscience and perfect reasoning, both in AI and in Philosophy (see [More97]). A particularly interesting suggestion was made by Hintikka in [Hint75], where he proposed the idea of considering *[logically] impossible [epistemically] possible worlds*; he seems not to have pursued this idea, though. This is the path followed in our work, as will be apparent in the rest of this document.

The main roots of these problems are the assumptions of *completeness* and *consistency* that underlie the possible worlds model. As McArthur notes in [McAr88], since worlds are *complete* the agent is forced to have beliefs about the way that everything would be in all of the accessible worlds; moreover, since worlds are *consistent* as well, everything that follows from the agent's beliefs must also be believed. Therefore, a natural solution to these problems could be reached by dropping these assumptions. This suggestion has been dismissed by most of the logicians and logically concerned philosophers of the Western tradition since Aristotle's time. Nevertheless, this possibility can be seriously entertained (some philosophers have indeed argued for the feasibility of this kind of worlds, *e.g.* Rescher and Brandom in [ReBr79]).

The *partiality* or *incompleteness* of possible worlds has been traditionally accepted in the AI literature, the most common justifications being 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], see [More97]).
- 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 a real computer.
- The agent can ignore some relevant rules (*e.g.* the agent may have not 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 (see [More97]).

Inconsistency is a totally different matter. Anyway, it can be argued in its favour with a number of ideas:

- The agent can be 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.
- It is indeed possible to define arguably interesting procedures of inquiry over inconsistent belief sets (as shown *e.g.* in [ReBr79]).
- These kinds of world are perfectly conceivable (and even depictable in pictures, as Escher proved so many times).
- Some theories have been considered acceptable for a certain period of time (*e.g.* Frege's set theory) for many people, but they have been proved to be inconsistent later (as Russell did with Frege's theory). This inconsistency does not mean that the theory has been useless and that all the work made on it has been wasted (many interesting insights can be gained even with the inconsistency proof).
- If a theory is expressed in first-order logic, it is not even decidable in general whether it is consistent or not, so these theories would be pretty useless if they could be used only in case they were previously proved to be consistent.
- A modified version of the analytic tableaux method is going to be used to perform the logical analysis of the agent's beliefs, as will be explained below, and the conception of possible worlds as (possibly) inconsistent sets induces a nice relationship between a tableau and the set of possible worlds that it represents (note that an open tableau can contain an inconsistent set of formulas, *e.g.* if the logical development of the tableau has not (yet) shown the contradiction that may be hidden in the set, waiting for further analysis to appear).

Therefore, we are going to avoid the classical logicians' reluctance towards (possibly) incomplete and (possibly) inconsistent possible worlds; they will be considered as (epistemologically and even ontologically) possible as the good old complete and consistent possible worlds. Consider the positive side of this move; if worlds are incomplete and inconsistent, both logical omniscience and perfect reasoning seem to vanish before our very eyes. The agent can clearly fail to believe some tautologies, and it does not have to believe any logical consequence of its beliefs. These facts will be easier to grasp later, when these worlds are used to model rational inquiry.

In fact, the expression *"possible world"* does not convey exactly the idea that we have of what a doxastic alternative is; in our framework, we will call them *"conceivable situations"*, rather than *"possible worlds"*. A *conceivable situation*, as its name suggests, is any situation that the modelled agent may conceive, irrespective of its partiality or its consistency. It may be a situation that it has experienced, that it has been told of, or even a situation that it has just imagined as possible. The only condition for an scenario to qualify as a conceivable situation is that the agent considers it so; it does not have to be either consistent or physically realizable. The main point is that a conceivable situation is *not* a model (in the logician's sense of the term). In the rest of the document the notion of conceivable situation will be considered as primitive, and will correspond to what the modelled agent considers as *"realities"*, be they experiential or just imagined.

2.2 Modelling the evolution of beliefs

Recall that the goal of this research is the definition of a general model of the process of rational inquiry, so something must be added to the classical possible worlds model (assuming that one indeed intends to keep the general conception of the model and the Kripke semantics). A very natural idea to model dynamic beliefs is to have some kind of *variability* in the main ingredients of the possible worlds model: the (assumed *fixed*) set of possible worlds W and the (also assumed *fixed*) accessibility relation R.

Imagine that the agent's beliefs in world w have to be analysed. This world is R-connected to worlds w_1,w_2 and w_3. A certain proposition `P` is true in w_2 and w_3, but not in w_1; therefore, `P` is not believed in w. If the set of worlds could be variable and not fixed, the following situation could arise: the agent could (for instance, as a consequence of an observation performed in the real world) notice that `P` holds indeed. This observation could be included in all the accessible worlds (w_1, w_2 and w_3), producing a modification of w_1 (which, in turn, would cause a modification in the agent's beliefs in w, which now would include P). This simple example shows that a variability in the contents of the possible worlds could account for experiences performed in reality, and could model the changes of beliefs produced by such observations. This is the kind of belief modification that will be used in the experimental dimension of analysis (see section 3.5).

Imagine that in the previous situation (P is true in w_2 and w_3, not in w_1, so P is not believed in w), a certain proposition Q is true in w_1 but false in w_2 and w_3 (therefore, it is not believed in w either). The agent, in the course of an inferential process performed on its beliefs, could reach the conclusion that Q is clearly unacceptable (*e.g.* it contradicts a large set of other actual beliefs). Therefore, it could conclude that the accessible worlds that contain Q are not viable alternatives to its present world, and thus they don't have to be considered accessible any more. This fact would imply that (w R w_1) would no longer hold, and that the set of doxastic alternatives to w would be reduced to $\{w_2, w_3\}$. But these two worlds contain P, and thus, *via* the standard Kripke semantics, the agent would now believe P in w. This example shows how a modification of the accessibility relation (*e.g.* a restriction in the set of possible doxastic alternatives) can indeed model a modification of the beliefs caused by an internal inferential process of the agent (other sources could have been considered; *e.g.* the agent could have observed the impossibility of Q as a result of an observation in its environment).

In summary, we think that the process of *rational inquiry* (including the consideration of observations made by the agent and the results of its own inferential processes) *can be modelled within the possible worlds model, keeping the Kripke semantics, but allowing dynamic changes in the set of possible worlds W and in the accessibility relation R.*

3 Multi-dimensional belief analysis

3.1 A syntactic representation of conceivable situations

A *conceivable situation* is (partially) represented by a set of propositional formulas in a 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 conceivable situation) and inconsistent (although perhaps the inconsistency is not apparent in the set, it may be hidden in its deductive closure). This characterization of conceivable situations implies that they can fail to contain some (even all) tautologies, and that the set of formulas that represents one of them does not have to be necessarily deductively closed. As mentioned above, this is a way to avoid the problems of logical omniscience and perfect reasoning.

In fact, it has been repeteadly proposed in the literature (see [More97]) to consider *impossible* possible worlds, in the sense of having possible worlds where the usual logical connectives do not behave in the usual way, or tautologies may not be true, or inconsistent formulas may hold. The drawback of these approaches is that, although they alleviate the problems of logical omniscience and perfect reasoning, they cause different problems. Basically, the main inconvenient of these approaches is that they are very limiting, in the sense that the agent has not got a minimal set of inference capabilities (*e.g.* it can believe P and (P $\Rightarrow$ Q) and not believe Q, or it can believe (P $\wedge$ Q) without believing either P or Q). In summary, there seems to be ([Halp86]) no way to make a knowledge-based analysis of the agent's beliefs (and also recall that our final aim is to model the evolution of beliefs over time, not a static set of beliefs in a certain point in time). The rest of the document defines a particular class of non-ideal agents, and shows how the evolution of their beliefs may be modelled in the framework of conceivable situations.

3.2 Multi-dimensional belief analysis in rational inquirers

In this document we are concerned with a special class of non-ideal agents, that will be called *rational inquirers*. These agents are permanently engaged in the process of *rational inquiry*; *i.e.* they keep trying to modify their beliefs, in order to make them as similar as possible to the facts in the real world. They try to get rid of those beliefs that do not reflect the state of the actual world, and to refine those that are discovered to be true. Thus, the belief set of this kind of agents is constantly being modified. Our aim is to give a formal model of the evolution of the beliefs of this kind of agents.

Rational inquirers try to make their beliefs as similar as possible to the actual world by performing a *multi-dimensional analysis of their beliefs*. The different dimensions of analysis that we consider are the following:

- A *logical* dimension, in which the agent will carry out some inferential (strictly deductive) processes.

- An *exploratory* dimension, in which the agent will be allowed to have *doubts*, to wonder whether it believes a given formula.
- An *experimental* dimension, in which the agent will perform tests in its environment, in order to obtain information that might be useful to refute doubtful beliefs or to confirm uncertain ones.

Moreover, other modifications of the set of beliefs will be considered; for instance, the agents's beliefs will also change in response to data received directly from the external environment, and the agent will be able to build an axiomatic model of its belief set. The following subsections describe the main ideas underlying each of these interwoven strands. The final section summarizes the document.

3.3 Logical dimension

3.3.1 Dynamic accessibility between conceivable situations

The agent's beliefs are expressed in a language $\mathcal{L}$, which is the language traditionally used in propositional modal logics of belief. This language is composed of a finite fixed set of basic propositions ($\mathcal{P}$), the basic logical operators ($\neg$, $\vee$, $\wedge$, $\Rightarrow$) and the modal doxastic operator B. $\mathcal{L}$ is restricted to standard propositional formulas prefixed by a (possibly empty) sequence of modal operators (*e.g.* $\neg(p \wedge (q \vee r))$, Bq, $BB(p \Rightarrow (\neg q \vee r))$). $\mathcal{L}_p$ is the strictly propositional part of $\mathcal{L}$.

In each conceivable situation (henceforth, a *cosi*) there are some propositions which are true and some propositions which the modelled agent believes. A *cosi* is (possibly partially) represented with a set of formulas of $\mathcal{L}_p$. Furthermore, no restriction is imposed on this set (it does not have to be consistent, it may be not deductively closed, it may fail to contain some tautologies, *etc.*). Intuitively, as explained above, a *cosi* can be envisioned as *any situation that the agent may consider as real*. It can be a situation that has happened to the agent, or maybe something it has imagined. It is useful to bear always in mind that they can be logically inconsistent or physically unrealizable, so they must not be confused with "possible worlds" or with logical (complete, consistent) models.

The standard Kripke semantics ([Krip63]) for modal logics of knowledge and belief is used throughout the logical analysis: a formula ϕ is believed by the agent in *cosi* w iff ϕ appears in all the doxastic alternatives to w (all the *cosis* related to w by the accessibility relation R). In the classical approaches the accessibility relation between "possible worlds" is predetermined, so the set of beliefs is constant. This fact implies that these approaches are unsuitable to model changing beliefs. In this new framework the aim is to model the evolution of the agent's beliefs over time. This evolution can be due to different facts, such as internal logical analysis of the agent or incorporation of data derived from tests performed in the environment in which the agent is located.

In the logical analysis the agent's set of beliefs in a *cosi* changes over time; this evolution will be formally modelled with a change in the accessibility relation between *cosis*, as suggested in section 2.2. In this model, a sequence of accessibility relations will be *generated*; each of them (*via* the standard Kripke semantics) will define a different set of beliefs. In the next section, we present a modified version of the analytic tableaux method, that is used by rational inquirers to perform a limited logical analysis of their beliefs.

3.3.2 Generation of the sequence of accessibility relations

The evolution of the agent's beliefs in a *cosi* will be described with an example. Assume that the agent's initial set of beliefs in a *cosi* w_e is the following:

$$Beliefs(w_e) \equiv \{\ B\mathtt{P},\ B(\mathtt{P} \vee \mathtt{Q}),\ B(\neg\mathtt{Q} \vee \mathtt{R}),\ BB\mathtt{P},\ BB(\neg\mathtt{Q} \vee \mathtt{R}),\ BB(\neg\mathtt{P} \vee \mathtt{Q})\ \}$$

The initial set of beliefs is just a set of formulas of $\mathcal{L}$; therefore, it is composed only by positive beliefs (however, in the course of the logical analysis the agent may attain negative beliefs, as will be shown later). Notice also that nested beliefs (at any level of nesting) may be considered.

The evolution of the agent's set of beliefs will be modelled with a sequence of accessibility relations. The initial accessibility relation, R_0, is obtained by applying the standard Kripke semantics in a *backwards* fashion; if the agent has a certain belief α, it will consider as doxastic alternatives all those *cosis* that contain α. If it believes that it believes a formula β, it will consider that it believes β in all its doxastic alternatives, and that means that it will consider all the *cosis* that contain β as doxastic alternatives of its doxastic alternatives, and so on. In the example used in this section, the following result would be obtained:

$$\forall w' \epsilon\, W,\ (w_e\ R_0\ w') \leftrightarrow \{\ \mathtt{P},\ (\mathtt{P} \vee \mathtt{Q}),\ (\neg\mathtt{Q} \vee \mathtt{R})\ \} \subseteq w'$$

Let's define $\overline{w_\alpha}$ as the class of all *cosis* that are R_0-accessible from w_e. The *cosis* R_0-accessible from (all the *cosis* in) $\overline{w_\alpha}$ would be those having the following property:

$$\forall w' \epsilon\ \overline{w_\alpha},\ \forall w''\ \epsilon\ W,\ (w'\ R_0\ w'') \leftrightarrow \{\ \mathtt{P},\ (\neg\mathtt{Q} \vee \mathtt{R}),\ (\neg\mathtt{P} \vee \mathtt{Q})\ \} \subseteq w''$$

A *cosi* w is R_0-accessible from w_e iff every proposition which is believed by the agent in w_e appears in w. A *cosi* w is R_0-accessible in two steps from w_e iff every proposition which is believed to be believed by the agent in w_e appears in w. The generation of the initial accessibility relation would continue until the deepest nested beliefs had been taken into account. The ontology of *cosis* just defined is represented in figure 1. In that figure each *cosi* w is represented by a rectangle that contains four columns, that have the following meaning (from left to right):

- α appears in the first column if $\alpha \epsilon w$ $(\alpha \epsilon \mathcal{L}_p)$
- α appears in the second column if $\alpha \not\epsilon w$ $(\alpha \epsilon \mathcal{L}_p)$
- α appears in the third column if α is believed in w
- α appears in the fourth column if α is not believed in w

The last two columns are included in an inner rectangle, to reinforce its doxastic interpretation. Classes of *cosis* ($\overline{w}$) are represented in the same way, and formulas are supposed to refer to every member of the class (*e.g.* the first column in $\overline{w_\alpha}$ shows the formulas that are included in all the *cosis* of that class).

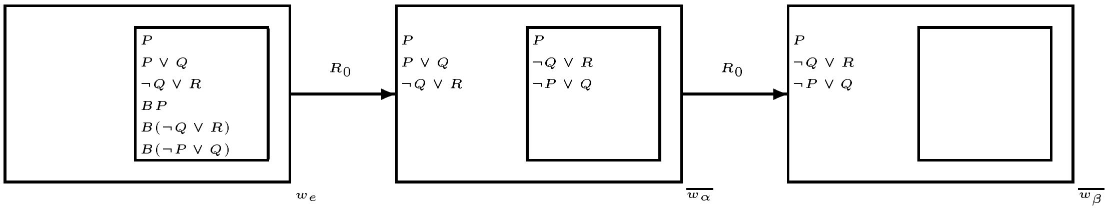

Figure 1: World w_e and its doxastic alternatives

w_e is R_0-related to all the *cosis* in class $\overline{w_\alpha}$. $\overline{w_\alpha} = \{\ w_{\alpha_1},\ w_{\alpha_2},\ \ldots\ \}$, where $w_{\alpha_i} = \{\mathtt{P},\ (\mathtt{P} \vee \mathtt{Q}),\ (\neg\mathtt{Q} \vee \mathtt{R})\} \cup \Delta_{\alpha_i}$. According to the above definitions, every proposition believed in w_e appears in all the *cosis* w_{α_1}, w_{α_2}, ... Each *cosi* in class $\overline{w_\alpha}$ is R_0-related to all the *cosis* in class $\overline{w_\beta}$; this class is formed by all those *cosis* that contain (at least) the formulas $\mathtt{P}$, $(\neg\mathtt{Q} \vee \mathtt{R})$ and $(\neg\mathtt{P} \vee \mathtt{Q})$.

The logical analysis of the agent's beliefs is performed using a modified version of the analytic tableaux method. The formulas contained in a tableau are divided into two sets, called *left column* and *right column*. Each tableau represents a set of *cosis*: *all those cosis that contain the formulas of its left column and do not contain any of the formulas of its right column*. The analysis may start *e.g.* with a tableau T_0 representing the class of *cosis* $\overline{w_\alpha}$, *i.e.* a tableau containing in its left column the formulas $\mathtt{P}$, $(\mathtt{P} \vee \mathtt{Q})$ and $(\neg\mathtt{Q} \vee \mathtt{R})$. Notice that T_0 not only represents the class of *cosis* $\overline{w_\alpha}$; it is also a representation of the initial first-level beliefs of the agent in w_e.

The splitting rule[1] of the analytic tableaux method has also been modified: 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 disjunction, namely that one (or both) of its members are true. The three subtableaux also keep all the formulas contained in T. The semantic meaning of the generation of these subtableaux is the following: the class of *cosis* represented by T is partitioned into four subclasses. These subclasses are formed taking into account the presence or absence of each of the two disjuncts of the analysed formula (*e.g.* if $(\alpha \vee \beta)$ is analysed, one of the subclasses that will be considered will contain all the *cosis* of the original class that contain α but do not contain β). The three subtableaux represent three of those four subclasses in which the original class is partitioned (the unrepresented subclass is composed by all those *cosis* of the original class that do not contain any of the disjuncts of the analysed formula).

[1]The logical analysis suggested in this report contains several rules for dealing with the logical operators in both columns of tableaux; nevertheless, in the example shown in this section the only rule that shall be needed is the one that analyses disjunctions contained in the left side of tableaux; this rule will be referred to as *the splitting rule*.

After the application of this rule to analyse (P ∨ Q) in T_0, the state shown in figure 2 is reached.

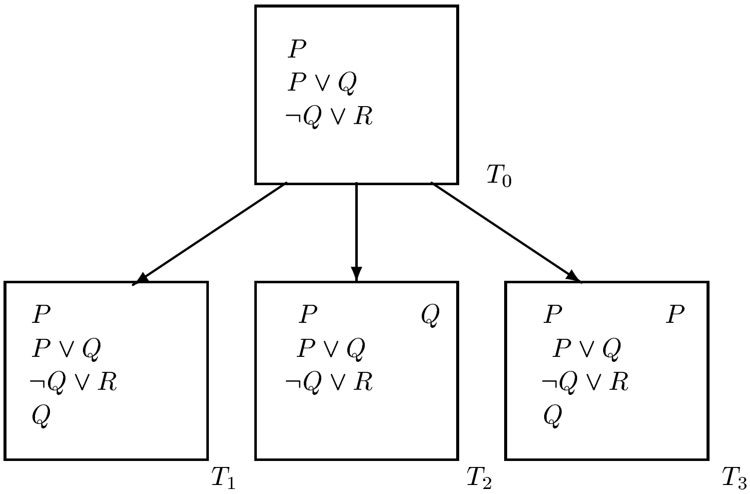

Figure 2: Splitting rule applied to (P ∨ Q) in T_0

One can wonder now which class of *cosis* is represented by T_1, T_2 or T_3. Recall that T_0 represents the class of *cosis* $\overline{w_\alpha}$. Consider the following partition of class $\overline{w_\alpha}$:

- $\overline{w_{\alpha_1}}$: *cosis* of $\overline{w_\alpha}$ that contain P and Q.
- $\overline{w_{\alpha_2}}$: *cosis* of $\overline{w_\alpha}$ that contain P but do not contain Q.
- $\overline{w_{\alpha_3}}$: *cosis* of $\overline{w_\alpha}$ that contain Q but do not contain P.
- $\overline{w_{\alpha_4}}$: *cosis* of $\overline{w_\alpha}$ that do not contain either P or Q.

It is clear that the tableau T_1 represents the class of *cosis* $\overline{w_{\alpha_1}}$, T_2 represents $\overline{w_{\alpha_2}}$ and T_3 represents $\overline{w_{\alpha_3}}$. The *cosis* 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 { P, (P ∨ Q), (¬Q ∨ R) } in which both P and Q fail to be true.

In the classical analytic tableaux method, those tableaux that contain an atom and its negation are *closed*, and thus dismissed from the logical analysis (because they cannot represent a model of the initial set of formulas, which is the aim of the method); the agent eliminates them from the analysis by *closing* them. In the tableaux analysis that is being considered in this document, the conditions that a tableau must satisfy in order to be eligible to be closed are the following:

- The tableau contains a formula and its negation in its left column.

  The tableau is representing a class composed by *cosis* that contain an explicit contradiction. They are considered as *logically impossible* and eliminated from the analysis by closing the tableau that represents them.

- The tableau contains the same formula both in its left and right columns.

  The tableau represents a class in which every *cosi* must contain and not contain the given formula, so it represents the empty set of *cosis* and can be dismissed from the analysis.

An important difference with the classical method is that a tableau is *not* closed automatically when one of these conditions holds. The agent *must* explicitly notice one of the above conditions and close it on purpose, after realising that the tableau represents an empty or a logically impossible class of *cosis* (*e.g.* a *cosi* may contain a lot of formulas and the agent may not have noticed that it contains a contradiction, even if it is as obvious as the presence of a formula and its negation). Note that the presence of contradictions does not prevent us from keeping our view of a tableau as the representation of a class of *cosis*, because, as explained above, *cosis* do not have to be logically consistent, physically realizable entities. They only have to be conceivable by the agent, but not *real*, in any wider sense of the word.

In the example, T_3 contains P both in its right and left columns, so it represents an empty set of *cosis*. If the agent decides to close it, the class of *cosis* represented by it is dismissed from the analysis. The remaining tableaux tree contains only T_0, T_1 and T_2. Thus, the effect of the analysis of the formula $(\mathtt{P} \vee \mathtt{Q})$ in T_0 is the restriction of the doxastic alternatives of w_e, that shrink from $\overline{w_\alpha}$ to its subclasses $\overline{w_{\alpha_1}}$ and $\overline{w_{\alpha_2}}$. The application of the splitting rule is modelled with the generation of a new accessibility relation, R_1, that restricts the set of R_0-accessible *cosis*. This restriction is shown in figure 3.

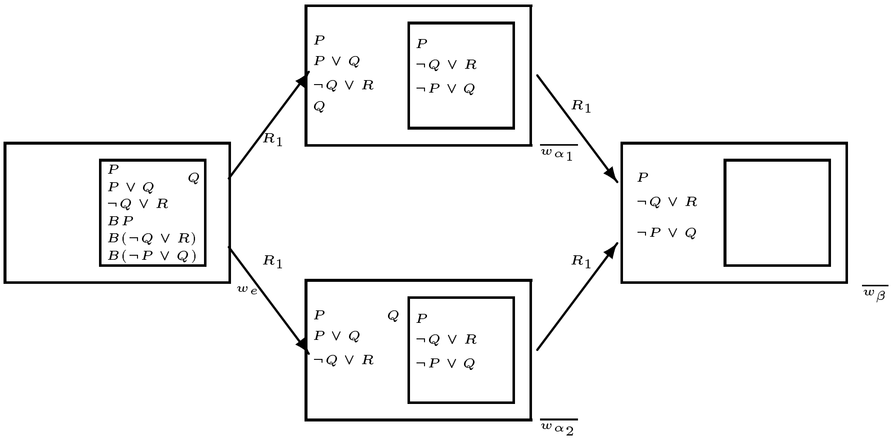

Figure 3: Generation of R_1

w_e was R_0-related to all the *cosis* in class $\overline{w_\alpha}$, but is R_1-related only to the ones in classes $\overline{w_{\alpha_1}}$ and $\overline{w_{\alpha_2}}$. The change of the accessibility relation implies a (possible) change in the agent's beliefs. The first-level beliefs in w_e at this point would still be $\{ \mathtt{P}, (\mathtt{P} \vee \mathtt{Q}), (\neg\mathtt{Q} \vee \mathtt{R}) \}$, because these are the only formulas common to both classes of doxastic alternatives. However, notice that the restriction of the doxastic alternatives has caused the addition of a new (negative) belief, $\neg B\mathtt{Q}$. This formula is a new belief because there are *cosis* accessible from w_e that do not contain Q (the ones in class $\overline{w_{\alpha_2}}$). This fact is reflected in figure 3, with the addition of Q in the fourth column of the representation of w_e. The second-level beliefs in w_e have not changed because the set of *cosis* accessible in two steps from w_e has not changed (it is still $\overline{w_\beta}$). Thus, the complete list of (positive and negative) beliefs held by the agent at this point in w_e would be the following: $\{ B\mathtt{P}, B(\mathtt{P} \vee \mathtt{Q}), B(\neg\mathtt{Q} \vee \mathtt{R}), \neg B\mathtt{Q}, BB\mathtt{P}, BB(\neg\mathtt{Q} \vee \mathtt{R}), BB(\neg\mathtt{P} \vee \mathtt{Q}) \}$.

If the agent wants to pursue the logical analysis, it can consider now *e.g.* the tableau that represents the class of worlds $\overline{w_\beta}$. This tableau (T_4) contains in its left column all the propositional formulas known to be included in all the *cosis* in this class, namely P, $(\neg\mathtt{Q} \vee \mathtt{R})$ and $(\neg\mathtt{P} \vee \mathtt{Q})$. Analysing the disjunction $(\neg\mathtt{Q} \vee \mathtt{R})$ in this tableau, the result shown in figure 4 is obtained.

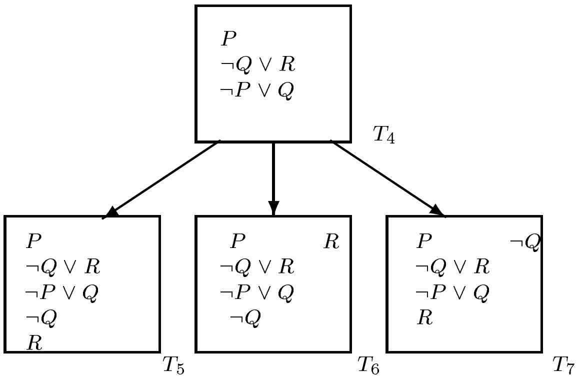

Figure 4: Analysis of $(\neg\mathtt{Q} \vee \mathtt{R})$ in T_4

The semantic counterpart of this syntactic analysis can now be considered. Recall that T_4 represents the class of *cosis* $\overline{w_\beta}$. This class can be partitioned into the following four subclasses:

- $\overline{w_{\beta_1}}$: *cosis* of $\overline{w_\beta}$ that contain $\neg\mathtt{Q}$ and $\mathtt{R}$.
- $\overline{w_{\beta_2}}$: *cosis* of $\overline{w_\beta}$ that contain $\neg\mathtt{Q}$ but do not contain $\mathtt{R}$.
- $\overline{w_{\beta_3}}$: *cosis* of $\overline{w_\beta}$ that contain $\mathtt{R}$ but do not contain $\neg\mathtt{Q}$.
- $\overline{w_{\beta_4}}$: *cosis* of $\overline{w_\beta}$ that do not contain either $\neg\mathtt{Q}$ or $\mathtt{R}$.

The tableaux generated in the last analysis (T_5, T_6 and T_7) represent the subclasses $\overline{w_{\beta_1}}$, $\overline{w_{\beta_2}}$ and $\overline{w_{\beta_3}}$, respectively. This analysis may be modelled with the generation of a new accessibility relation, R_2, in which accessibility has been restricted again, because the *cosis* in class $\overline{w_{\beta_4}}$ are no longer accessible from the ones in classes $\overline{w_{\alpha_1}}$ and $\overline{w_{\alpha_2}}$, as they were with R_0 and R_1. This restriction is shown in figure 5.

The agent would now notice that there are *cosis* accessible from those in classes $\overline{w_{\alpha_1}}$ and $\overline{w_{\alpha_2}}$ that do not contain $\neg Q$ (the *cosis* belonging to class $\overline{w_{\beta_3}}$), so this formula would not be believed in any of the *cosis* of these classes. The same situation happens with $\mathtt{R}$, which is not contained in any of the *cosis* in class $\overline{w_{\beta_2}}$ and therefore is not believed in the previous level either. This situation is shown in figure 5, where $\neg B\neg\mathtt{Q}$ and $\neg B\mathtt{R}$ appear in the first level (in the fourth column of both $\overline{w_{\alpha_1}}$ and $\overline{w_{\alpha_2}}$). The information in the first level is transmited to w_e by the standard Kripkean procedure; as all the doxastic alternatives of this *cosi* contain $\neg B\neg\mathtt{Q}$ and $\neg B\mathtt{R}$, these formulas must be believed by the agent in w_e, as shown in figure 5.

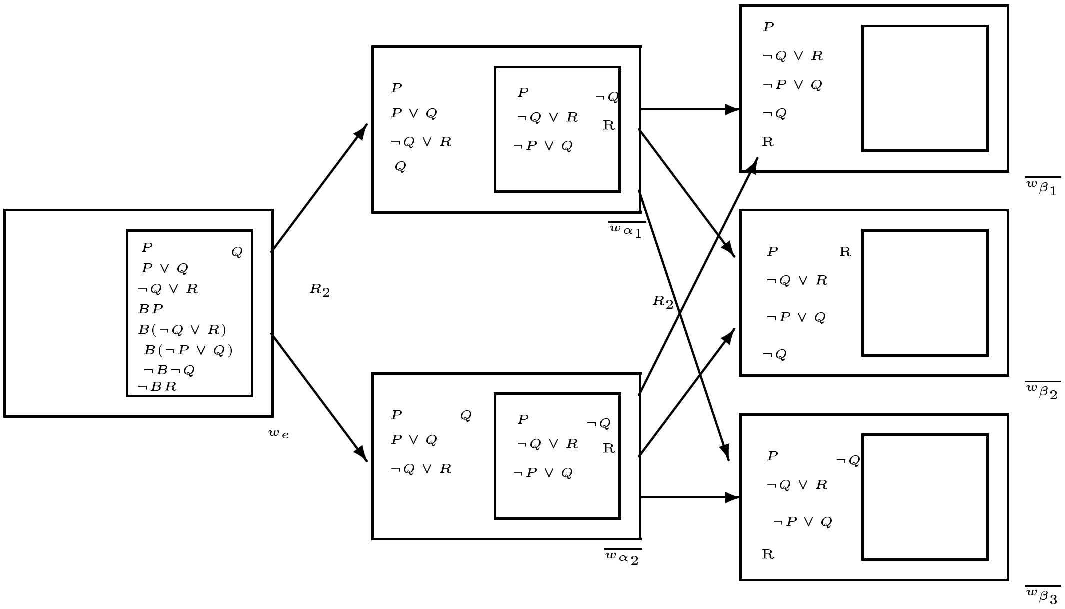

Figure 5: Generation of R_2

The main idea in the model of the logical analysis is the following: to keep a (direct or indirect) access to all those classes of *cosis* represented by all open tableaux. There is a tableaux tree to be considered for each level of nesting of the initial set of beliefs. When a disjunction is analysed in an open tableau, a new accessibility relation is generated. In this new accessibility relation, there is no longer access to the class of *cosis* represented by that tableau, but to some subclasses of that class of *cosis* (those represented by the generated subtableaux that have not been closed). In the following list you can find a brief summary of the relationship between the syntactic tableaux analysis and its semantic counterpart (*cosis* and the accessibility relation between them, that define the agent's beliefs via the Kripke semantics):

- Tableau $\equiv$ Class of *cosis*.
- Closed tableau $\equiv$ Special type of tableau, that is discovered by the agent to represent either an empty set of *cosis* or a class of (logically impossible) *cosis*.
- Application of a splitting rule $\equiv$ Generation of a new accessibility relation between *cosis*.

Following the example, the agent's beliefs at this point are $\{\ B\mathtt{P},\ B(\mathtt{P}\vee\mathtt{Q}),\ B(\neg\mathtt{Q}\vee\mathtt{R}),\ BB\mathtt{P},\ BB(\neg\mathtt{Q}\vee\mathtt{R}),\ BB(\neg\mathtt{P}\vee\mathtt{Q}),\ B\neg B\neg\mathtt{Q},\ B\neg B\mathtt{R},\ \neg B\mathtt{Q}\ \}$. The agent may now decide to proceed with the logical analysis, because there are still open tableaux that contain formulas that have not been analysed yet. The five possibilities of analysis at this point are:

- The formula $(\neg\mathtt{Q}\vee\mathtt{R})$ may be analysed in T_1 (that represents class $\overline{w_{\alpha_1}}$) and T_2 ($\overline{w_{\alpha_2}}$).
- The formula $(\neg\mathtt{P}\vee\mathtt{Q})$ may be analysed in T_5 ($\overline{w_{\beta_1}}$), T_6 ($\overline{w_{\beta_2}}$) and T_7 ($\overline{w_{\beta_3}}$).

Let's assume that the agent decides to analyse $(\neg\mathtt{P}\vee\mathtt{Q})$ in T_5. The result of this analysis is shown in figure 6.

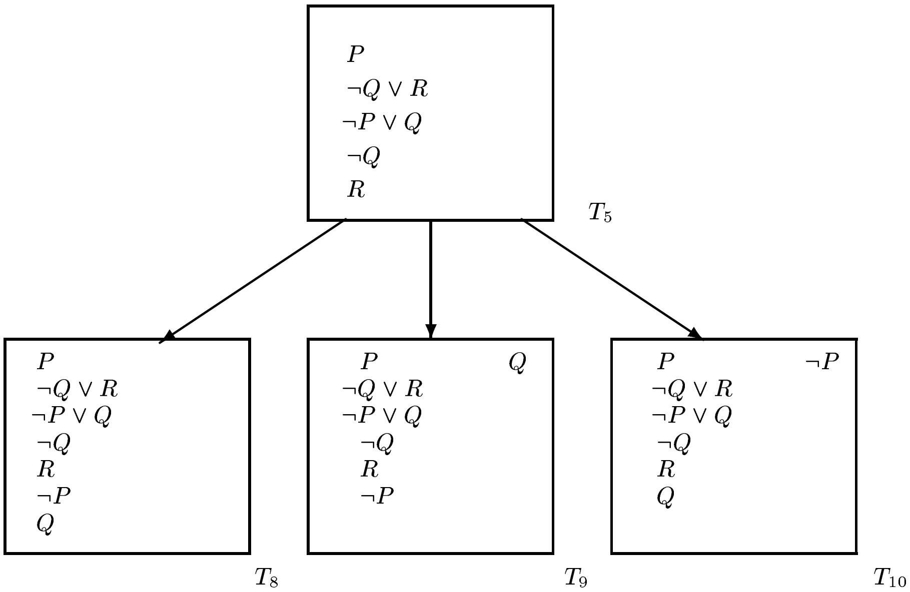

Figure 6: Analysis of $(\neg\mathtt{P}\vee\mathtt{Q})$ in T_5

As usual, $\overline{w_{\beta_1}}$ can be partitioned into four subclasses, taking into account the absence or presence of $\neg\mathtt{P}$ and $\mathtt{Q}$. Three of these subclasses are represented by the three subtableaux generated in the analysis of $(\neg\mathtt{P}\vee\mathtt{Q})$ (T_8, T_9 and T_{10}), while the fourth one is not even taken into account (that would be the class of all those *cosis* in $\overline{w_{\beta_1}}$ that contain the disjunction but do not contain any of the disjuncts).

It can be noticed that T_8 and T_9 contain both $\mathtt{P}$ and $\neg\mathtt{P}$ in their left columns, and therefore the classes of *cosis* represented by these tableaux may be considered by the agent as logically impossible and dismissed from the analysis by closing those tableaux. T_{10} contains $\mathtt{Q}$ and $\neg\mathtt{Q}$ in its left column, so it may also be closed and the class of (logically impossible) *cosis* that it represents is no longer taken into account. This argument shows that all the subtableaux generated in this analysis are closed; that means that the agent has no longer access to the *cosis* of class $\overline{w_{\beta_1}}$, which were those *cosis* represented by T_5. In summary, the agent has explored all the subclasses of $\overline{w_{\beta_1}}$ and has discovered that none of them is logically possible. Thus, in our model we would now generate a new accessibility relation, R_3, shown in figure 7, in which that class of *cosis* is no longer accessible. It can also be noticed that this elimination does not imply any change in the positive and negative beliefs of previous levels. The positive beliefs in the first level are still $\mathtt{P}$, $(\neg\mathtt{Q}\vee\mathtt{R})$ and $(\neg\mathtt{P}\vee\mathtt{Q})$, because these are still the only formulas known to appear in all its doxastic alternatives (the *cosis* in classes $\overline{w_{\beta_2}}$ and $\overline{w_{\beta_3}}$); the negative beliefs in the first level do not change either ($\neg\mathtt{Q}$ and $\mathtt{R}$), because for each of these formulas there are accessible *cosis* that do not contain it ($\neg\mathtt{Q}$ is not contained in $\overline{w_{\beta_3}}$ and $\mathtt{R}$ is not contained in $\overline{w_{\beta_2}}$). If the situation has not changed in classes $\overline{w_{\alpha_1}}$ and $\overline{w_{\alpha_2}}$, it cannot have changed in w_e either, because these two classes contain all those *cosis* that are accessible from w_e.

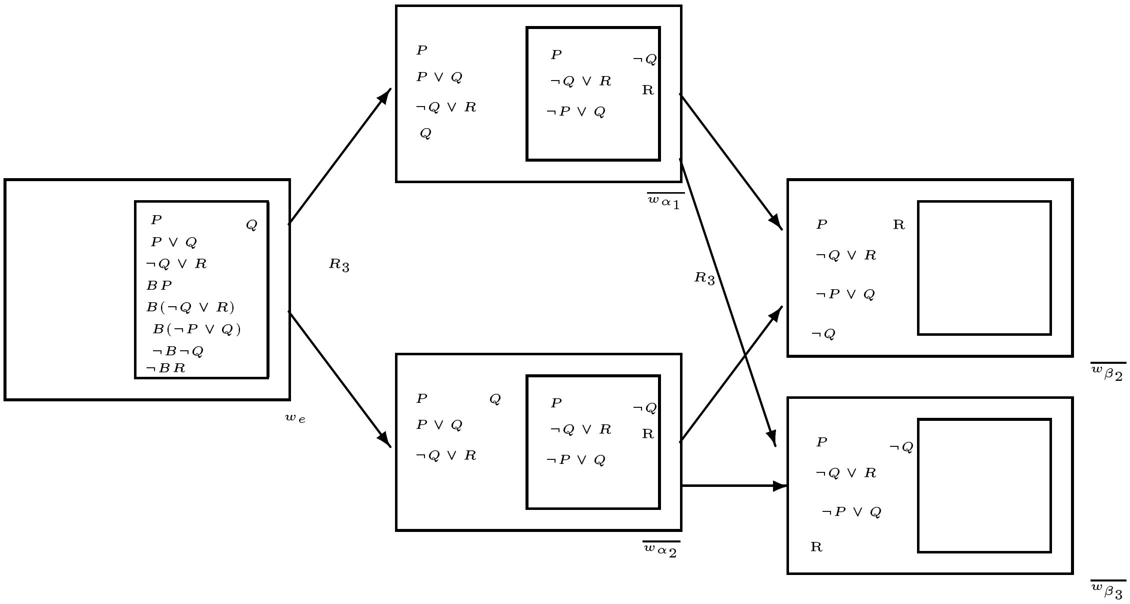

Figure 7: Generation of R_3

A similar situation arises when the agent analyses ($\neg$P $\vee$ Q) in T_6, which is the tableau that represents the class of *cosis* $\overline{w_{\beta_2}}$ (see figure 8), because the three subtableaux generated in this analysis also contain contradictions in their left columns, so all of them may be closed (T_{11} and T_{12} contain P and $\neg$P, whereas T_{13} contains Q and $\neg$Q).

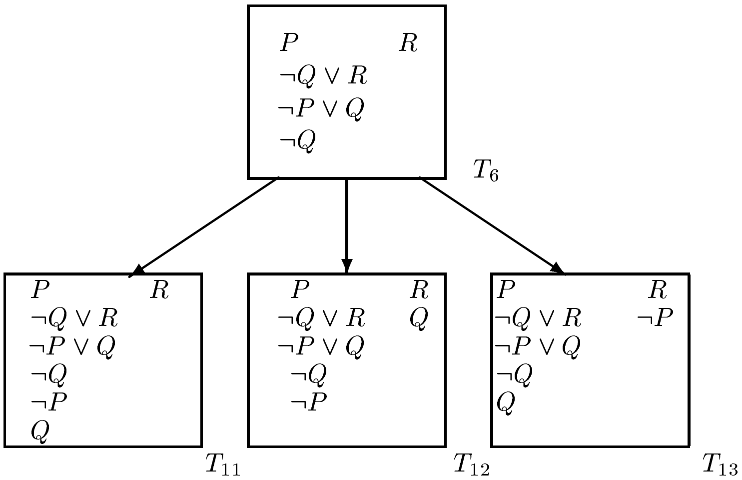

Figure 8: Analysis of ($\neg$P $\vee$ Q) in T_6

Recall that T_6 is the representation of the class of *cosis* $\overline{w_{\beta_2}}$. As T_{11}, T_{12} and T_{13} are generated from the analysis of a formula in T_6, they represent subclasses of $\overline{w_{\beta_2}}$. All of these subtableaux are closed, so all these subclasses are considered as logically impossible and eliminated from the analysis. That means that the agent has no longer access to any *cosi* in class $\overline{w_{\beta_2}}$, as shown in figure 9, that represents the accessibility relation generated as a consequence of the last tableau analysis, R_4.

The change of the accessibility relation from R_3 to R_4 models a change in the agent's beliefs. Now the only *cosis* accessible from $\overline{w_{\alpha_1}}$ and $\overline{w_{\alpha_2}}$ are those in class $\overline{w_{\beta_3}}$. All of these *cosis* contain P, ($\neg$Q $\vee$ R),

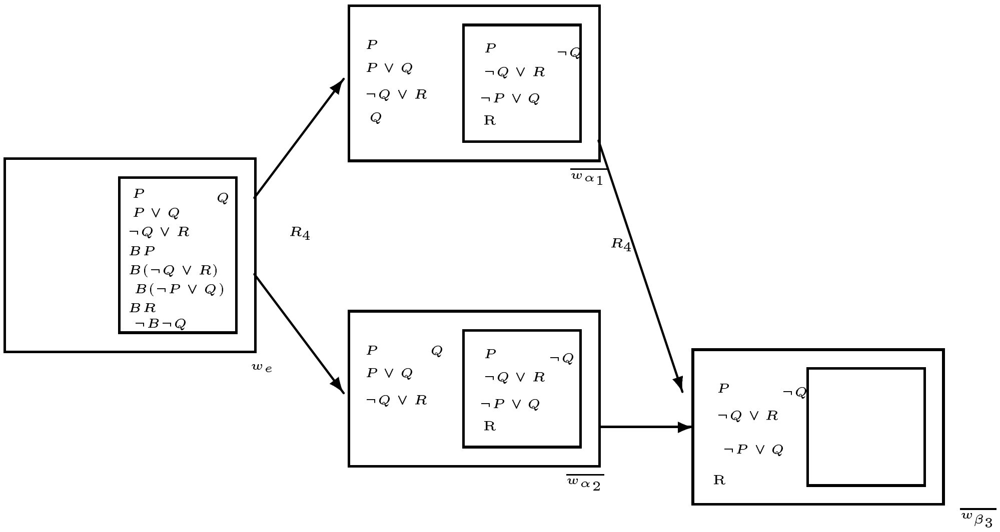

Figure 9: Generation of R_4

($\neg$P $\vee$ Q) and R, and therefore these four formulas are believed in all the *cosis* contained in the classes $\overline{w_{\alpha_1}}$ and $\overline{w_{\alpha_2}}$. The first three formulas were already believed in all of these *cosis* from the beginning of the analysis, but notice that R was explicitly *not* believed in those *cosis* when the previous accessibility relation, R_3, was considered. It is easy to see that if a formula is positively believed at some point of the analysis, it will continue to be considered as a positive belief in the rest of the analysis, whereas if a formula is a negative belief at some point of the analysis, it may be the case that it is transformed into a positive belief at a later stage of the logical analysis. This result is due to the fact that the sequence of accessibility relations is increasingly restrictive. If a formula is a positive belief, it means that it is true in all doxastic alternatives; no matter how these alternatives are restricted, the formula will still be true in the remaining accessible *cosis*. If a formula is a negative belief, that means that there are accessible *cosis* which do not contain that formula; if the set of accessible *cosis* is restricted in such a way that these *cosis* are eliminated (as has happened in the example), the formula will be contained in all remaining doxastic alternatives and it will be considered as a positive belief.

Now the agent may proceed with the analysis of ($\neg$P $\vee$ Q) in T_7 (the tableau that represents the class of *cosis* $\overline{w_{\beta_3}}$). The result of the analysis of that disjunction in shown in figure 10.

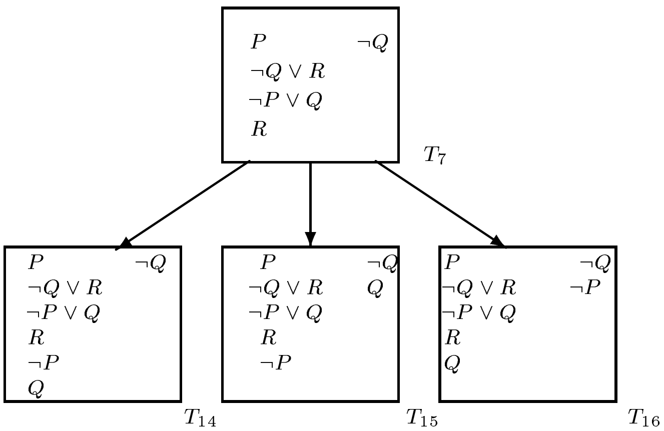

Figure 10: Analysis of ($\neg$P $\vee$ Q) in T_7

Three subtableaux are generated in the analysis of $(\neg P \vee Q)$ in T_7. These subtableaux (T_{14}, T_{15} and T_{16}) represent three subclasses of the class of *cosis* $\overline{w_{\beta_3}}$. The first two subtableaux may be closed because they contain P and ¬P in their left columns, and therefore they represent classes of (logically impossible) *cosis*. The remaining tableau, T_{16}, represents a class of *cosis* ($\overline{w_{\beta_{3_3}}}$) that is defined as follows:

- $\overline{w_{\beta_{3_3}}}$: all those worlds of class $\overline{w_{\beta_3}}$ that contain Q and do not contain ¬P.

As usual, the semantic counterpart of the syntactic logical analysis is the generation of a new accessibility relation between *cosis*, R_5, in which the previous accessibility relation is restricted. In this case, all those *cosis* that had access (through R_4) to the *cosis* in class $\overline{w_{\beta_3}}$ are only R_5-related to the ones in class $\overline{w_{\beta_{3_3}}}$. This situation is depicted in figure 11.

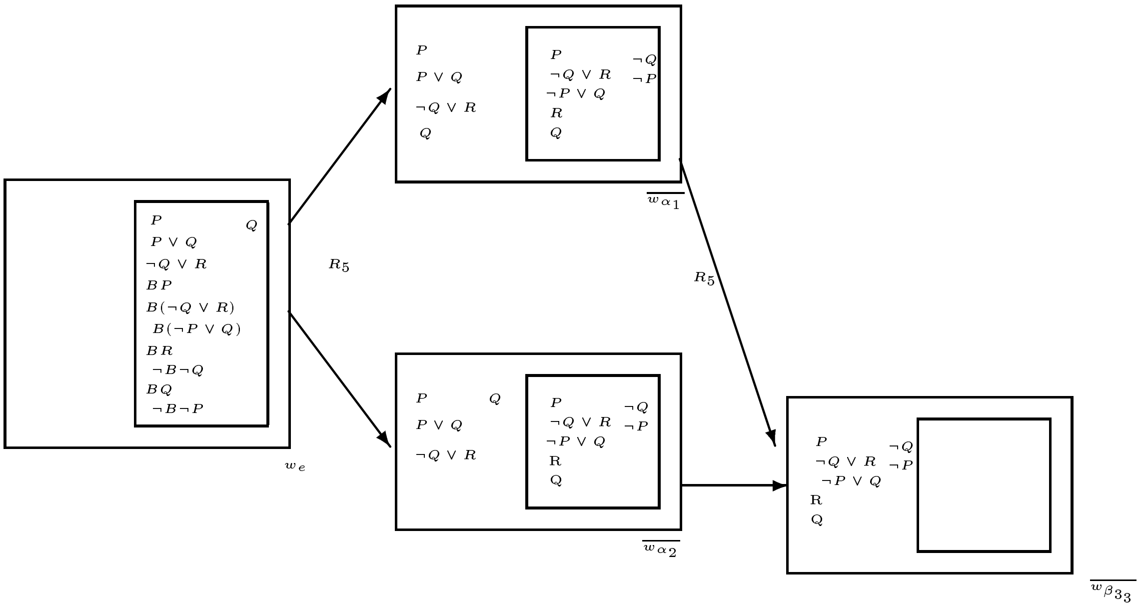

Figure 11: Generation of R_5

After the generation of this new accessibility relation, the agent's beliefs are updated by applying the standard Kripke semantics. The beliefs at $\overline{w_{\alpha_1}}$ and $\overline{w_{\alpha_2}}$ are all the formulas that are included in all the *cosis* in $\overline{w_{\beta_{3_3}}}$, because this is the only class of *cosis* R_5-accessible from those two classes. Therefore, the beliefs in these classes are P, (¬Q ∨ R), (¬P ∨ Q), R and Q, as shown in the third column of their representations. Moreover, the agent also knows that ¬Q and ¬P are not included in $\overline{w_{\beta_{3_3}}}$, and thus these two formulas are not believed in $\overline{w_{\alpha_1}}$ and $\overline{w_{\alpha_2}}$. This fact is shown in the fourth column of the representation of these classes in figure 11. Now the agent's (second level) beliefs in w_e may be updated, by applying again the Kripke semantics. The result would be the addition of two new beliefs:

- BQ is believed in w_e because Q is believed in all its doxastic alternatives; that means that Q is contained in all the doxastic alternatives of the doxastic alternatives of w_e; in short, all the *cosis* R_5-accessible from w_e in two steps contain Q.
- $\neg B$¬P is believed in w_e because ¬P is not believed in any of its doxastic alternatives; that means that, for each *cosi* w R_5-accessible from w_e, there exists a *cosi* w' R_5-accessible from w such that w' does not contain ¬P.

There are still two open tableaux with formulas that have not yet been analysed; thus, the agent may decide to proceed with the logical analysis by analysing another disjunction. For instance, it may choose to analyse (¬Q ∨ R) in T_1, as shown in figure 12.

T_1 represents $\overline{w_{\alpha_1}}$, so each of the subtableaux generated in the analysis of (¬Q ∨ R) represents a subclass of this class. T_{17} and T_{18} may be closed because they represent classes of logically impossible *cosis*, as they have both Q and ¬Q in their left columns. On the other hand T_{19} is not closed; we may

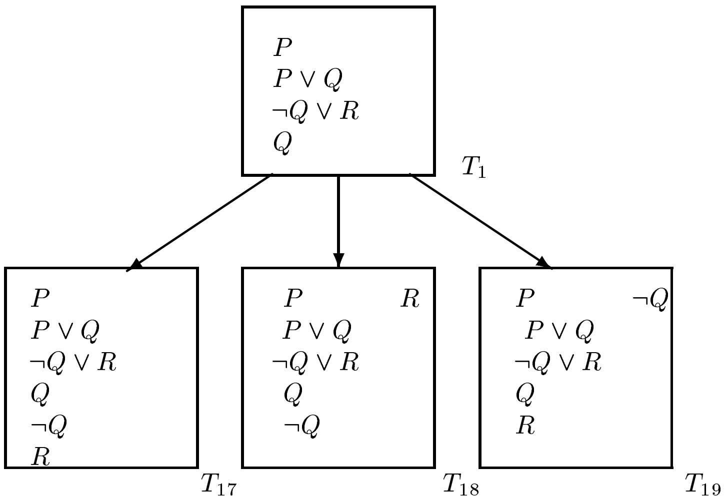

Figure 12: Analysis of ($\neg$`Q` $\vee$ `R`) in T_1

model this fact with the construction of a new accessibility relation, R_6, in which all those *cosis* that had access to $\overline{w_{\alpha_1}}$ will have access only to the subclass of this class represented by T_{19}, which shall be called $\overline{w_{\alpha_{1_3}}}$. This subclass is defined as follows:

- $\overline{w_{\alpha_{1_3}}}$: all those *cosis* of class $\overline{w_{\alpha_1}}$ that contain `R` and do not contain $\neg$`Q`.

This new accessibility relation, R_6, is represented in figure 13.

As usual, the agent updates its beliefs after this restriction in the set of accessible *cosis*. The positive beliefs in w_e do not change, because the set of formulas included in all its doxastic alternatives (all the *cosis* in classes $\overline{w_{\alpha_{1_3}}}$ and $\overline{w_{\alpha_2}}$) is still the same (`P`, (`P` $\vee$ `Q`) and ($\neg$`Q` $\vee$ `R`)). However, the set of negative beliefs changes, because there are accessible *cosis* (those in $\overline{w_{\alpha_{1_3}}}$) that do not contain $\neg$`Q`, so the agent does not believe $\neg$`Q` in w_e, as reflected in the fourth column of its representation in figure 13.

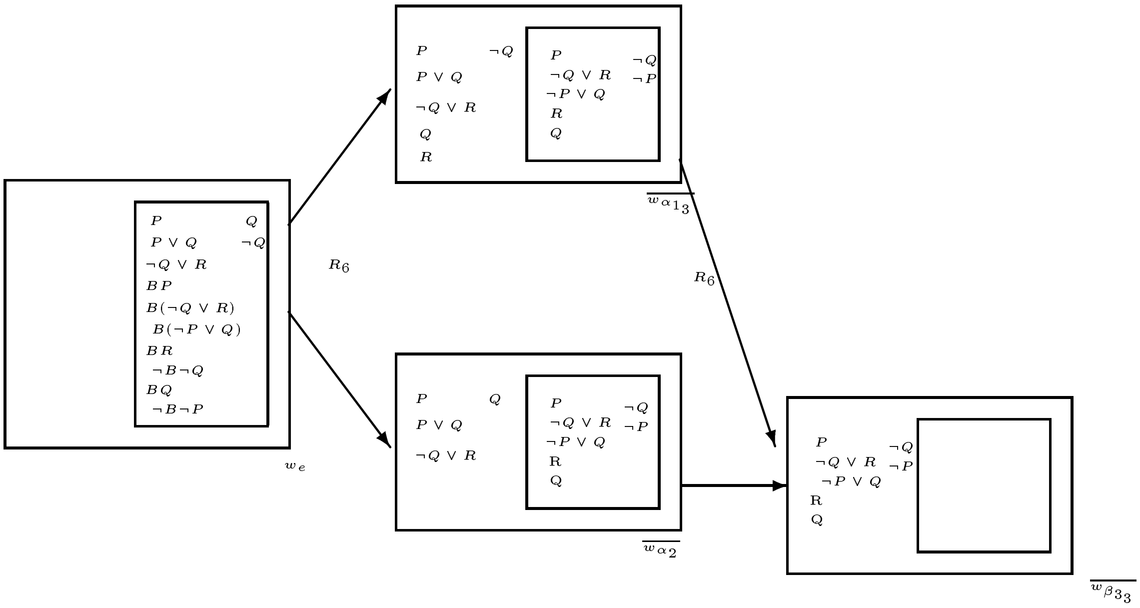

Figure 13: Generation of R_6

There is only one formula left to be analysed, $(\neg\mathtt{Q} \vee \mathtt{R})$ in T_2, the tableau that represents the class of *cosis* $\overline{w_{\alpha_2}}$. The analysis of this formula is shown in figure 14.

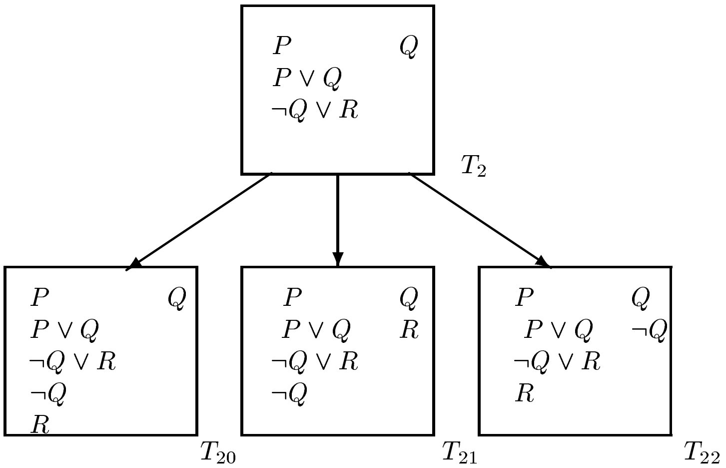

Figure 14: Analysis of $(\neg\mathtt{Q} \vee \mathtt{R})$ in T_2

The subtableaux generated in this analysis, T_{20}, T_{21} and T_{22}, represent subclasses of the class of *cosis* represented by T_2, $\overline{w_{\alpha_2}}$. This class can be partitioned into the following subclasses:

- $\overline{w_{\alpha_{2_1}}}$: *cosis* of $\overline{w_{\alpha_2}}$ that contain $\neg\mathtt{Q}$ and $\mathtt{R}$.
- $\overline{w_{\alpha_{2_2}}}$: *cosis* of $\overline{w_{\alpha_2}}$ that contain $\neg\mathtt{Q}$ but do not contain $\mathtt{R}$.
- $\overline{w_{\alpha_{2_3}}}$: *cosis* of $\overline{w_{\alpha_2}}$ that contain $\mathtt{R}$ but do not contain $\neg\mathtt{Q}$.
- $\overline{w_{\alpha_{2_4}}}$: *cosis* of $\overline{w_{\alpha_2}}$ that do not contain either $\neg\mathtt{Q}$ or $\mathtt{R}$.

T_{20} represents $\overline{w_{\alpha_{2_1}}}$, T_{21} represents $\overline{w_{\alpha_{2_2}}}$ and T_{22} represents $\overline{w_{\alpha_{2_3}}}$. Therefore, the semantic counterpart of the application of the splitting rule is the generation of a new accessibility relation between *cosis*, R_7, in which the *cosis* that had access to class $\overline{w_{\alpha_2}}$ (w_e) now have access only to its subclasses $\overline{w_{\alpha_{2_1}}}$, $\overline{w_{\alpha_{2_2}}}$ and $\overline{w_{\alpha_{2_3}}}$. The restriction of the accessibility relation is shown in figure 15.

At this point the set of w_e's doxastic alternatives has changed, and the agent's beliefs in this *cosi* may be updated by applying the standard Kripke semantics. The second-level beliefs have not changed, because the set of *cosis* R_7-accessible from w_e in two steps is the same than the set of *cosis* R_6-accessible from w_e in two steps $(\overline{w_{\beta_{3_3}}})$. The first-level positive beliefs in w_e have not changed either, because the only formulas that appear in all its doxastic alternatives are still $\mathtt{P}$, $(\mathtt{P} \vee \mathtt{Q})$ and $(\neg\mathtt{Q} \vee \mathtt{R})$. However, the set of negative beliefs has changed; notice that a formula α is defined to be a negative belief at a *cosi* w $(\neg\mathrm{B}\alpha)$ iff there is at least one doxastic alternative to w that does not contain α. With this definition in mind, it is easy to check that the agent's negative beliefs in w_e at this point of the analysis are the formulas $\mathtt{Q}$ (which is not contained in $\overline{w_{\alpha_{2_1}}}$, $\overline{w_{\alpha_{2_2}}}$ and $\overline{w_{\alpha_{2_3}}}$), $\neg\mathtt{Q}$ (which is not contained in $\overline{w_{\alpha_{1_3}}}$ and $\overline{w_{\alpha_{2_3}}}$) and (a new negative belief) $\mathtt{R}$. This formula is not believed in w_e because the agent has discovered in the logical analysis that there are *cosis* accessible from w_e (those in class $\overline{w_{\alpha_{2_2}}}$) that do not contain it.

All formulas in all open tableaux have already been analysed, so this would be the end of the purely logical analysis of the initial set of beliefs. Therefore, the final beliefs of the agent in *cosi* w_e are: $\{B\mathtt{P}$, $B(\mathtt{P} \vee \mathtt{Q})$, $B(\neg\mathtt{Q} \vee \mathtt{R})$, $BB\mathtt{P}$, $BB(\neg\mathtt{Q} \vee \mathtt{R})$, $BB(\neg\mathtt{P} \vee \mathtt{Q})$, $BB\mathtt{R}$, $BB\mathtt{Q}$, $B\neg B\neg\mathtt{Q}$, $B\neg B\neg\mathtt{P}$, $\neg B\mathtt{Q}$, $\neg B\neg\mathtt{Q}$, $\neg B\mathtt{R}\}$.

In summary, the logical analysis causes the evolution of the agent's beliefs due to internal deductive procedures. This evolution is modelled with the generation of a sequence of accessibility relations, R_0, R_1, R_2, …, in which every accessibility relation is more restrictive than the previous one (and thus, the agent's beliefs increase in this process).

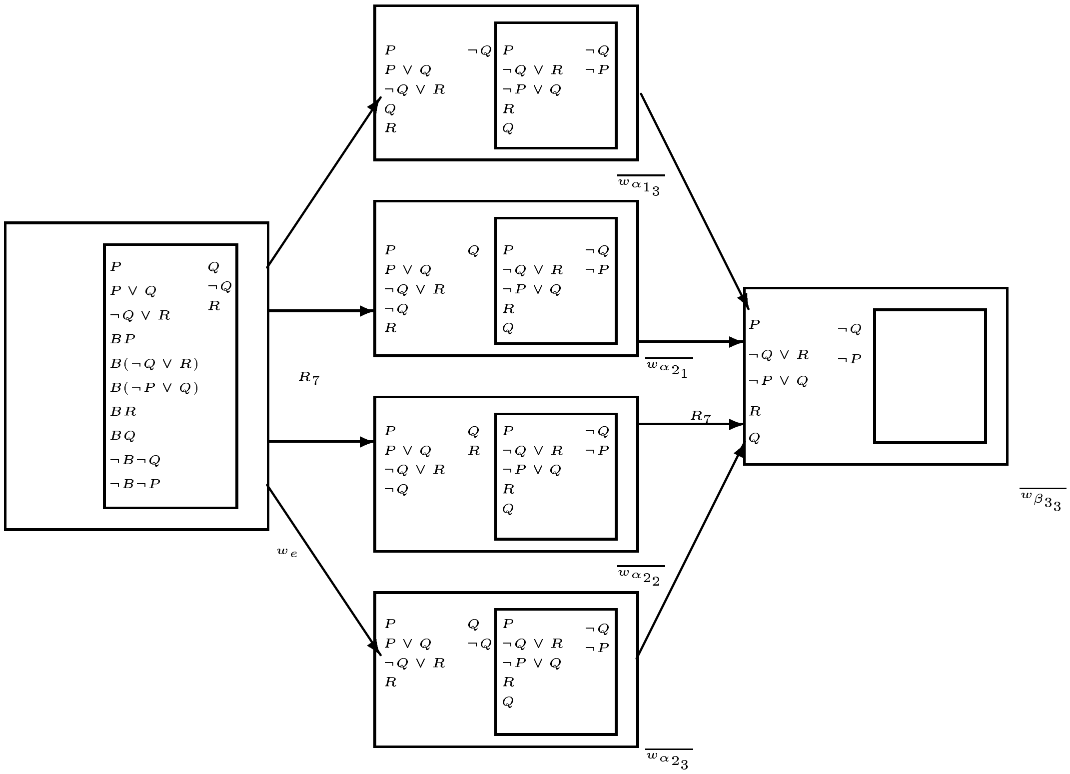

Figure 15: Generation of R_7

3.4 Exploratory dimension

In the logical analysis of beliefs the classical analytic tableaux method has been modified in a number of ways. There is another important difference with respect to the classical method: *the agent is not allowed to add to an open tableau any tautology*. If tautologies could be freely added into the tableaux, the problem of logical omniscience could not be avoided. This prohibition is based on the idea that a tableau is just a partial representation of a set of *cosis*, and it cannot be modified with no apparent reason.

In the exploratory dimension of analysis, the agent is allowed to introduce some formulas in the open tableaux; more specifically, the agent may pose itself questions, may introduce *doubts* in the analysis, may wonder whether a certain formula ϕ is or not the case. Technically, this idea is implemented by allowing the introduction of instances of the *Axiom of the Excluded Middle* $(\phi \vee \neg\phi)$ in the open tableaux of the logical analysis. The use of this particular tautology seems a natural way to allow the agent to have *doubts*, to wonder whether it believes some formula (ϕ) or its negation $(\neg\phi)$. This exception permits the introduction of the formula $(\phi \vee \neg\phi)$ in a tableau, which is later split into two subtableaux containing ϕ in one column and $\neg\phi$ in the other. 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.

The introduction of self-posed questions (by the agent, *via* Excluded Middle) and the consequent splitting of tableaux, suggest a simple explanation of two hitherto unrelated fenomena. One is purely logical: the closing of all tableaux generated by a set of statements (representing *e.g.* one's beliefs) now means, simply, that all the conceivable situations potentially contained in (or derived from) the given set are impossible; the immediate consequence is that the agent, after having explored all the open possibilities, ceases to believe in the set (a process discovered with much fracas by the Pythagoreans, and usually called *reductio ad absurdum*). The second lies more in the province of epistemologists or even philosophers of science: suppose (considering predicate logic for a second) that the agent doubts of the validity of a general law such as *All birds fly* (*i.e.* $\forall x\ (Bird(x) \Rightarrow Flies(x))$). It adds to its set of beliefs (in the exploratory dimension of analysis) the formula $\forall x(Bird(x) \Rightarrow Flies(x)) \vee \neg\forall x(Bird(x) \Rightarrow Flies(x))$ and, when it analyses this disjunction (in the logical dimension), it will have access to two classes of conceivable situations: in the first class the law will hold, while the second will contain $\neg\forall x(Bird(x) \Rightarrow Flies(x))$, from which by standard (first-order) tableaux processing the agent would get $Bird(a)$ and $\neg Flies(a)$ (for some (undetermined) a). At this point, the agent can notice that it can increase its beliefs if it can dismiss this class of *cosis*; if the agent looks *actively* in the real world for an individual a such that $Bird(a)$ and $\neg Flies(a)$ and it cannot find it, then it can conclude that an individual with these properties does not exist; when this conclusion is finally reached, the agent can declare this class of *cosis* as *physically impossible*, and so the agent will have grounds not to consider it any more as a conceivable, realizable alternative. Then the tableau that represents this class of *cosis* would be closed and dismissed from the analysis. At that moment all the *cosis* accessible by the agent would contain $\forall x(Bird(x) \Rightarrow Flies(x))$ so, by definition, the agent would believe this law. This is an example of how the agent may combine different dimensions of analysis in order to refine its set of beliefs.

3.5 Experimental dimension

In the experimental dimension of analysis the agent is allowed to make experiments in its environment and to incorporate the resulting propositions (externally obtained, as opposed to the internally obtained propositions of the logical analysis previously described) in all open tableaux of the logical analysis. Thus, this dimension models the agent's acquisition of data in the actual world (obtained *via* experiences, tests, *etc.*). The root of this dimension can be traced as far as Kant, who argued in his *Critique of pure reason* that Reason has to take into account observations of the environment; he thinks that Reason must not approach Nature as a student, that takes everything that its teacher chooses to say, but as a judge who formulates questions and compels the witnesses to answer them. Nature has to be constrained to give answers to questions of Reason's own determining.

This dimension is very much related to the logical dimension. The logical analysis can guide the experimental analysis, suggesting which experiences or tests the agent can perform in the actual world to gain knowledge. The agent may find out whether its actual set of beliefs is accurate by trying to confirm its predictions by experience.

3.6 External inputs

We also allow the agent to incorporate into the analysis information that it receives from the environment. These new pieces of information are added to the analysis by introducing them in the open tableaux of the logical analysis. This incorporation may cause a tableau to be closed, if it fulfills any of the tableau closing conditions and the agent realizes that fact. Thus, a tableau may be closed by different reasons:

- As a result of the logical tableaux analysis, the tableau may describe a contradictory class of *cosis* or an empty class of *cosis*; if the agent notes this fact, it will close that tableau (a purely *logical closing*) and it will cease to take that class of *cosis* into account in the rest of the analysis.
- As a result of information received from the environment (as a direct external input or in the experimental dimension of analysis), one of the closing conditions may hold in a tableau; thus, if the agent realizes that situation, it will close the tableau (a logical closing due to empirical reasons).
- The agent may fail to find in the real world a set of individuals that satisfy the properties stated in a tableau. Thus, it may consider the class of *cosis* represented by the tableau as *physically impossible* and it may close that tableau in order to eliminate them from further analysis.

One important issue arises here, namely the difference between monotonic and non-monotonic beliefs. The former are the ones obtained from the purely logical closing of tableaux, while the latter are those derived from the *empirical* closings of tableaux, caused by the other dimensions of analysis, that may be defeated later (*e.g.* the agent can later notice that a fact (received directly from another agent) was not true, or that the environment's answer to a question, in the experimental dimension, was not correct, or that there indeed exists an individual with a desired set of properties). This means that this work will have connections with the AI field of *belief revision*, which has received much attention in recent years.

3.7 Limited belief analysis

There are several aspects of the belief analysis that limit the reasoning capabilities of *rational inquirers*, and prevent them from being *ideal agents*:

- Tableaux are not closed automatically in the logical analysis. The agent must explicitly note the appearence of one of the closing conditions before closing the tableau.
- Tautologies may not be freely added to tableaux (with the exception of the instances of the Axiom of the Excluded Middle that may be used in the exploratory dimension of analysis, see section 3.4).
- 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* analysing 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 analyse any tableau. In this way we can model imperfect reasoning due to a lack of resources, whereas with the previous differences 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, or to pose itself a question in the exploratory analysis).

3.8 Building an axiomatic system

The fact that the agent's beliefs are not deductively closed can be interpreted in this way: they are actually "deductively closed", but not according to normal logic but to some "local" (the agents's own) inference rules. The main idea would be to represent the beliefs of the agent with an axiomatic system (a set of core axioms plus a set of inference rules), in a Konolige's *deduction model of belief* fashion ([Kono86]). The agent's beliefs would then be all those propositions that can be obtained from the set of axioms with the application of the inference rules.

Notice that this dimension does not perform any deduction on the set of beliefs (as the logical dimension does), but merely transforms the set of beliefs into another form, namely an axiomatic system.

The main motivation of this dimension is a concept that could be called *cognitive economy*. There are several advantages of an axiomatic system *vs.* a set of beliefs, *e.g.* the following:

- Ease of manipulation and representation.
- Representation of different levels of acceptance of the beliefs (*e.g.* the easier it is to obtain a proposition from the axioms, the easier it is for the agent to believe it).
- Compactness of the axiomatic system.
- Different viewpoints on the same set of beliefs (with different choices of the axioms and inference rules).

4 Summary

- The main aim of this work is to provide a way of modelling the process of rational inquiry, unlike the standard approach of modelling an agent's belief set in a certain fixed point in time.
- This process of rational inquiry is modelled within the possible worlds framework and its classical Kripke semantics. To avoid the problems of logical omniscience and perfect reasoning that appear in this semantics, two important changes to the classical approaches are made:
  - We consider *conceivable situations* (*cosis*) as the primitive semantic entities; they include all the situations that the modelled agent is capable of imagining or considering, regardless of their partiality or inconsistency. A *cosi* is partially represented with a set of propositional formulas.
  - Both the set of *cosis* (W) and the accessibility relation between them (R) are not constant, fixed, but variable. This variability accounts for the evolution of the agent's beliefs over time (through the usual Kripke semantics).
- We define a class of non-ideal rational agents, called *rational inquirers*, that perform a multi-dimensional dynamic analysis of their beliefs, in order to make them as similar as possible to the facts in the real world. This analysis includes:
  - A logical dimension, in which the agent makes some (limited) deductive inferences in its belief set, using a modified version of the analytic tableaux method.

  - An experimental dimension, that allows the agent to take into account all the data that it can obtain from the actual world. The agent performs tests in the real world in order to increase its knowledge. The tests to be performed are suggested by the logical analysis (linking in a novel fashion the rational and empirical components of the classical notion of inquiry). The agent modifies its beliefs if it finds a contradiction between a logical derivation and experimental data.
  - An exploratory dimension, where the agent is allowed to have doubts, to wonder whether a formula is true or not.

- Moreover, we also consider some additions to this analysis:
  - The agent may receive data from the environment and incorporate this information into its belief set.
  - An axiomatic component, which transforms the belief set into a more compact format, namely into an axiomatic set. This idea does not appear in the classical views of inquiry in the philosophical literature.

5 References

[CRW97] Cavedon, L., Rao, A., Wobcke, W. (Eds.), "**Intelligent Agent Systems: Theoretical and Practical Issues**", LNAI 1209, Springer Verlag 1997.

[FaHa85] Fagin, R., Halpern, J., "**Belief, awareness and limited reasoning**", *Procs. of IJCAI-85*, pp. 491-501, 1985.

[Halp86] Halpern, J., "**Reasoning about knowledge: an overview**", *Proceedings of TARK-86*, Ed. J. Halpern, pp. 1-17, 1986.

[HaMo92] Halpern, J., Moses Y., "**A guide to completeness and complexity for modal logics of knowledge and belief**", *Artificial Intelligence* 54, pp. 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, pp. 475-484, 1975.

[Kono86] Konolige, K., "**A Deduction Model of Belief**", Morgan Kaufmann, San Mateo, CA, 1986.

[Krip63] Kripke, S., "**A semantical analysis of modal logic I: normal modal propositional calculi**", *Zeitschrift für Mathematische Logik und Grundlagen Mathematik* 9, pp. 67-96, 1963.

[McAr88] McArthur, G., "**Reasoning about knowledge and belief: a survey**", *Computational Intelligence* 4, pp. 223-243, 1988.

[More97] Moreno, A., "**How to avoid knowing it all**", Research Report DEI-RR-97-013, Computer Science Department, University Rovira i Virgili (URV), 1997.

[MoSa96a] Moreno, A., Sales, T., "**Dynamic belief analysis**", *Workshop on Agent Theories, Architectures and Languages, ATAL-96*, at *ECAI-96*, Budapest, August 1996 (see [MoSa97a]).

[MoSa96b] Moreno, A., Sales, T., "**Limited logical belief analysis**", *Workshop on Theoretical and Practical Foundations of Intelligent Agents*, at *PRICAI-96*, Cairns, Australia, 1996 (see [MoSa97b]).

[MoSa97a] Moreno, A., Sales, T., "**Dynamic belief analysis**", in [MWJ97], pp. 87-102, 1997.

[MoSa97b] Moreno, A., Sales, T., "**Limited logical belief analysis**", in [CRW97], pp. 104-118, 1997.

[MWJ97] Müller, J., Wooldridge, M., Jennings, N. (Eds.), "**Intelligent Agents III: Agent theories, architectures and languages**", LNAI 1193, Springer Verlag 1997.

[Perl94] Perlis, D., "**Logic for a lifetime**", CS-TR-3278, Computer Sci. Dept., Univ. of Maryland, 1994.

[ReBr79] Rescher, N., Brandom, R., "**The Logic of Inconsistency**", Rowman and Littlefield Eds., 1979.

[RuNo95] Russell, S., Norvig, P., "**Artificial Intelligence. A Modern Approach.**", Prentice Hall Series in Artificial Intelligence, 1995.

[Shoh93] Shoham, Y., "**Agent-oriented programming**", *Artificial Intelligence* 60, pp. 51-92, 1993.