Generalizing AGM to a Multi-agent Setting

Guillaume Aucher,
Université Paul Sabatier – University of Otago
IRIT – Équipe LILaC
118 route de Narbonne
F-31062 Toulouse cedex 9 (France)
E-mail: aucher@irit.fr

Abstract

We generalize AGM belief revision theory to the multi-agent case. To do so, we first generalize the semantics of the single-agent case, based on the notion of interpretation, to the multi-agent case. Then we show that, thanks to the shape of our new semantics, all the results of the AGM framework transfer. Afterwards we investigate some postulates that are specific to our multi-agent setting. Finally, we give an example of revision operator that fulfills one of these new postulates and give an example of revision on a concrete example.

Keywords: Belief revision, epistemic logic, private announcement

1 Introduction

AGM belief revision theory [1] has been designed for a single agent. A natural idea is to extend it to the multi-agent case. As in AGM, we consider the beliefs of *one* agent, that we call Y (like You). But in this case this agent, in her representation of the surrounding world, will have to deal not only with facts about the world but also with the other agents' perception of the surrounding world. So, we will have to extend or generalize the single agent semantics in order to take into account this multi-agent aspect.

Besides, in a multi-agent setting, we have to be careful about what kind of multi-agent belief revision we study and consequently about the nature of events we consider. In this paper we are interested in *private announcements* made to Y. A private announcement is an event where Y learns privately (from an external source for example) some piece of information about the original situation, the other agents not being aware of anything. This piece of information might be factual or epistemic, i.e. about some other agents' beliefs. Finally, by *private multi-agent belief revision*, we mean the revision that Y must perform in case the private announcement of ϕ made to her contradicts her beliefs. So far, this kind of revision has not been studied.

In the case of private announcement, the other agents' beliefs clearly do not change. For example, suppose *you* (Y) believe p, and agent j believes p (and perhaps even that p is common belief of Y and j). When a third external agent privately tells you that $\neg p$ then j still believes p and you still believe that j believes p (and that j believes that p is common belief). This static aspect of private announcements is similar to the static aspect of AGM belief revision in a single-agent case: in both cases the world does not change but only agent Y's beliefs about the world change. So, it is reasonable to expect that the AGM framework can be extended to private multi-agent belief revision. In this paper we propose a natural generalization. The central device will be internal models, of which AGM models are a particular case.

The paper is organized as follows. In Section 2, we recall belief revision theory in the line of [14]. In Section 3, we first introduce the notions of multi-agent possible worlds and internal models in order to adequately represent agent Y's perception of the surrounding world. We then propose an equivalent representation and we generalize the AGM framework to the multi-agent case. In Section 4, we investigate some additional rationality postulates specific to our multi-agent approach. Finally in Section 5, we give an example of a revision operator and an application of this operator to a concrete example.

Note 1.1
All the proofs of this paper can be found in the appendix.

2 The single agent case: the AGM approach

In this paper Φ is a *finite* set of propositional letters and L the propositional language defined over Φ. Often, the epistemic state of the agent is represented by a belief set K. This belief set is an infinite set of propositional formulas closed under logical consequence and whose formulas represent the beliefs of some agent, here called Y. However, we prefer to represent epistemic states by finite belief bases as it is easier to handle by computers. For that, we follow the approach of [14].

As argued by Katsuno and Mendelzon, because Φ is finite, a belief set K can be equivalently represented by a propositional formula ψ: $K = Cn(\psi) = \{\chi \mid \psi \to \chi\}$. So $\chi \in K$ iff $\psi \to \chi$. Now, given a belief base ψ and a sentence ϕ, $\psi \circ \phi$ denotes the revision of ψ by ϕ; that is the new belief base obtained by adding ϕ to the old belief base ψ and giving up some formulas if necessary to keep consistency. In fact, given a revision operator $*$ on belief sets, one can define a corresponding operator $\circ$ on belief bases as follows: $\psi \circ \phi \to \chi$ iff $\chi \in Cn(\psi) * \phi$. Thanks to this correspondence, Katsuno and Mendelzon set some rationality postulates for this revision operator $\circ$ on belief bases which are equivalent to the AGM rationality postulates for the revision $*$ on belief sets. These postulates express how a rational agent should revise her belief set when she receives incoming information that she believes to be true.

Remark 2.1
Here are the AGM rationality postulates:

(K*1) $K * \phi$ is a belief set
(K*2) $\phi \in K * \phi$
(K*3) $K * \phi \subseteq K + \phi$
(K*4) If $\neg\phi \notin K$ then $K + \phi \subseteq K * \phi$
(K*5) $K * \phi = K_\perp$ iff ϕ is unsatisfiable
(K*6) If $\phi \leftrightarrow \phi'$ then $K * \phi = K * \phi'$
(K*7) $K * (\phi \wedge \phi') \subseteq (K * \phi) + \phi'$
(K*8) If $\neg\phi' \notin K * \phi$ then $(K * \phi) + \phi' \subseteq K * (\phi \wedge \phi')$

LEMMA 2.2
[14] Let * be a revision operator on belief sets and $\circ$ its corresponding operator on belief bases. Then * satisfies the 8 AGM postulates $(K*1)-(K*8)$ iff $\circ$ satisfies the postulates $(R1)-(R6)$ below:

(R1) $\psi \circ \phi \rightarrow \phi$.
(R2) if $\psi \wedge \phi$ is satisfiable, then $\psi \circ \phi \leftrightarrow \psi \wedge \phi$.
(R3) If ϕ is satisfiable, then $\psi \circ \phi$ is also satisfiable.
(R4) If $\psi_1 \leftrightarrow \psi_2$ and $\phi_1 \leftrightarrow \phi_2$, then $\psi_1 \circ \phi_1 \leftrightarrow \psi_2 \circ \phi_2$.
(R5) $(\psi \circ \phi) \wedge \phi' \rightarrow \psi \circ (\phi \wedge \phi')$.
(R6) If $(\psi \circ \phi) \wedge \phi'$ is satisfiable, then $\psi \circ (\phi \wedge \phi') \rightarrow (\psi \circ \phi) \wedge \phi'$.

So far our presentation to revision was syntactic. Now we are going to give the semantics of AGM revision and then set some links between the two.

Let $\mathcal{I}$ be the set of all interpretations of the finite propositional language L. $Mod(\psi)$ denotes the set of all interpretations that make ψ true. Let $\mathcal{M}$ be a set of interpretations of L. $form(\mathcal{M})$ denotes a formula whose set of models is equal to $\mathcal{M}$.

A *pre-order* $\leq$ over $\mathcal{I}$ is a reflexive and transitive relation on $\mathcal{I}$. A pre-order is *total* if for every $I, J \in \mathcal{I}$, either $I \leq J$ or $J \leq I$. Consider a function that assigns to each propositional formula ψ a pre-order $\leq_\psi$ over $\mathcal{I}$. We say this assignment is *faithful* if the following three conditions hold:

1. If $I, I' \in Mod(\psi)$, then $I <_\psi I'$ does not hold.
2. If $I \in Mod(\psi)$ and $I' \notin Mod(\psi)$, then $I <_\psi I'$ holds.
3. If $\psi \leftrightarrow \psi'$, then $\leq_\psi = \leq_{\psi'}$.

Let $\mathcal{M}$ be a subset of $\mathcal{I}$. An interpretation I is minimal in $\mathcal{M}$ with respect to $\leq_\psi$ if $I \in \mathcal{M}$ and there is no $I' \in \mathcal{M}$ such that $I' <_\psi I$. Let

$$Min(\mathcal{M}, \leq_\psi) = \{I \mid I \text{ is minimal in } \mathcal{M} \text{ with respect to } \leq_\psi\}.$$

The following (representation) theorem shows that a revision operator satisfying the above rationality postulates, and which is therefore a syntactic object, can be represented semantically by a faithful assignment.

THEOREM 2.3 (Representation theorem)
[14] Revision operator $\circ$ satisfies postulates $(R1)-(R6)$ iff there exists a faithful assignment that maps each belief base ψ to a total pre-order $\leq_\psi$ such that $Mod(\psi \circ \phi) = Min(Mod(\phi), \leq_\psi)$.

PROOF. The detailed proof can be found in [14], we just give a sketch here. The "if" direction is straightforward. For the "only-if" direction, the key is the definition of a faithful assignment for each belief base in terms of $\circ$. For any interpretations I and I' ($I = I'$ is permitted), we define a relation $\leq_\psi$ as: $I \leq_\psi I'$ iff either $I \in Mod(\psi)$ or $I \in Mod(\psi \circ form(I, I'))$. ∎

This semantic revision process is described in Figure 1. In this figure, the dots represent interpretations and the diagonal line separates the interpretations satisfying ϕ from the interpretations satisfying $\neg\phi$. The interpretations in the inner circle are the worlds that satisfy ψ and thus correspond to $Mod(\psi)$. The other circles represent the ordering $\leq_\psi$: if $w <_\psi w'$ then w is within a smaller circle than w' and if $w =_\psi w'$ then w and w' are in between the same successive circles. So the farther an interpretation is from the inner circle, the farther it is from ψ. The interpretations in the hatched part are then the interpretations that satisfy ϕ and which are the closest to ψ. Therefore they represent $Mod(\psi \circ \phi) = Min(Mod(\phi), \leq_\psi)$.

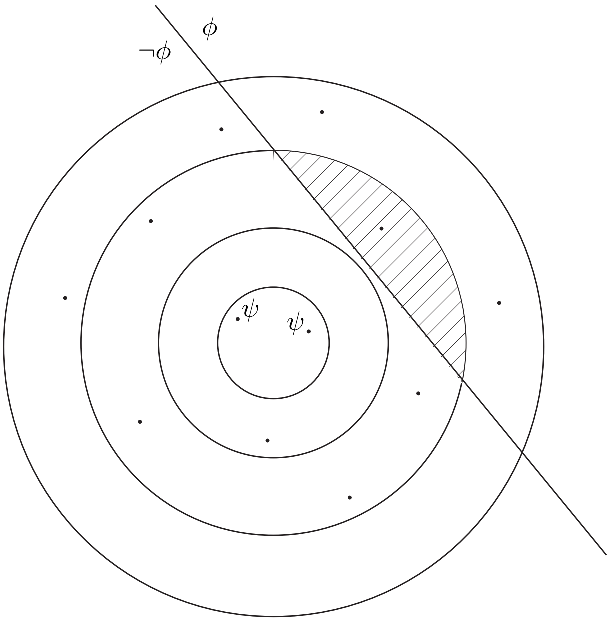

Fig. 1. AGM belief revision

Grove proposed another semantic approach based on a system of spheres [11]. But one can show that his framework can be recast in the one just described.

3 The multi-agent case

3.1 Some preliminaries

In this paper, G is a fixed set of agents such that $Y \in G$.

3.1.1 Epistemic logic

We first recall the basics of epistemic logic [12, 9]. An epistemic model M is a tuple $M = (W, \{R_j \mid j \in G\}, val)$ where W is a set of worlds, R_j are accessibility relations indexed by agents $j \in G$ and val is a function that assigns to each $w \in W$ a subset of Φ. We define $R_j(w)$ by $R_j(w) = \{v \mid wR_jv\}$ and $|M|$ is the number of worlds in M. Finally, a $KD45_G$ epistemic model is an epistemic model whose accessibility relations are serial, transitive and euclidean.

Classically, an epistemic model M is given with an actual world w_a: (M, w_a).

Intuitively, a (pointed) epistemic model (M, w_a) represents from an external point of view how the actual world w_a is perceived by the agents G. The possible worlds W are the relevant worlds needed to define such a representation and the valuation val specifies which propositional facts (such as 'it is raining') are true in these worlds. Finally the accessibility relations R_j model the notion of belief. We set $w' \in R_j(w)$ in case in world w, agent j considers the world w' (epistemically) possible.

Finally, the submodel of M generated by a set of worlds $S \subseteq M$ is the restriction[1] of M to the worlds $\{(\bigcup_{j \in G} R_j)^*(w); w \in S\}$ (where $(\bigcup_{j \in G} R_j)^*$ is the reflexive transitive closure of $(\bigcup_{j \in G} R_j)$, see [7] for details). In case the submodel of M generated by a set of worlds $S \subseteq M$ is M itself then M is said to be generated by S. Intuitively, the submodel of M generated by a set of worlds S contains all the relevant information in M about these worlds S.

Now we can define a language for epistemic models which will enable us to express things about them.

$$\mathcal{L}^C : \phi := \top \mid p \mid \neg\phi \mid \phi \wedge \phi \mid B_j\phi \mid C_{G_1}\phi,$$

where j ranges over G, G_1 over subsets of G, and p over Φ. We also define $\mathcal{L}$ as the sub-language of $\mathcal{L}^C$ without common belief operator C_{G_1}. The semantics of these two languages are defined as usual as follows.

$$\begin{array}{lll}
M, w \models \top & & \\
M, w \models p & \text{iff} & w \in V(p) \\
M, w \models \neg\phi & \text{iff} & \text{not } M, w \models \phi \\
M, w \models \phi \wedge \phi' & \text{iff} & M, w \models \phi \text{ and } M, w \models \phi' \\
M, w \models B_j\phi & \text{iff} & \text{for all } v \in R_j(w), M, v \models \phi \\
M, w \models C_{G_1}\phi & \text{iff} & \text{for all } v \in (\bigcup_{j \in G_1} R_j)^+(w)\ M, v \models \phi
\end{array}$$

where $(\bigcup_{j \in G_1} R_j)^+$ is the transitive closure of $\bigcup_{j \in G_1} R_j$.

So agent j believes ϕ in world w (formally $M, w \models B_j\phi$) if ϕ is true in all the worlds that the agent j considers possible (in world w). For example, in the pointed epistemic model (M, w) of Figure 5, agent Y does not know whether p is true or not: $M, w \models \neg B_Y p \wedge \neg B_Y \neg p$. Agent Y also believes that A does not know neither: $M, w \models B_Y(\neg B_A p \wedge \neg B_A \neg p)$. Finally, agent Y believes that A believes that she does not know whether p is true or not: $M, w \models B_Y B_A(\neg B_Y \neg p \wedge \neg B_Y p)$.

3.1.2 Bisimulation

We now recall the definition of a bisimulation.

Definition 3.1

Let Z be a non-empty relation between worlds of two finite epistemic models $M = (W, \{R_j \mid j \in G\}, val)$ and $M' = (W', \{R'_j \mid j \in G\}, val')$. We define the property of Z being a bisimulation in w and w', noted $Z : M, w \leftrightarrows M', w'$ as follows.

[1] Let $M = (W, \{R_j \mid j \in G\}, val)$ be an epistemic model. The restriction of M to a set of worlds S is the submodel $M' = (W', \{R'_j \mid j \in G\}, val')$ of M defined as follows. $W' = W \cap S$; $R'_j := R_j \cap (S \times S)$ for all $j \in G$; and $val'(w) = val(w)$ for all $w \in W'$.

1. If wZw' then $val(w) = val'(w')$;
2. if wZw' and $v \in R_j(w)$ then there exists $v' \in R_j(w')$ such that vZv';
3. if wZw' and $v' \in R_j(w')$ then there exists $v \in R_j(w)$ such that vZv'.

We say that M, w and M', w' are bisimilar, noted $M, w \leftrightarrows M', w'$, iff there is a relation Z such that $Z : M, w \leftrightarrows M', w'$. It can be shown (in case M and M' are finite) that $M, w \leftrightarrows M', w'$ iff for all $\phi \in \mathcal{L}^C$, $M, w \models \phi$ iff $M', w' \models \phi$. So, intuitively, two epistemic models are bisimilar if they contain the same information.

3.1.3 Characterization of finite models

Finally, we will also use the following proposition.

Proposition 3.2
[6][4][17] Let M be a *finite* epistemic model and $w \in M$. Then there is an epistemic formula $\delta_M(w) \in \mathcal{L}^C$ (involving common knowledge) such that

1. $M, w \models \delta_M(w)$;
2. for every finite epistemic model M' and world $w' \in M'$, if $M', w' \models \delta_M(w)$ then $M, w \leftrightarrows M', w'$.

This proposition tells us that a finite epistemic model can be completely characterized (modulo bisimulation) by an epistemic formula. For example, the pointed epistemic model (M, w) in Figure 5 is characterized by the following epistemic formula: $\delta_M(w) = p \wedge (\neg B_Y p \wedge \neg B_Y \neg p) \wedge (\neg B_A p \wedge \neg B_A \neg p) \wedge C_G((\neg B_Y p \wedge \neg B_Y \neg p) \wedge (\neg B_A p \wedge \neg B_A \neg p))$.

This proposition will be very useful to prove that the results of the single agent case of AGM belief revision theory transfer to the multi-agent case.[2]

3.2 From possible world to multi-agent possible world

3.2.1 The notion of multi-agent possible world

In the AGM framework, one considers a single agent Y. The possible worlds introduced are supposed to represent how the agent Y perceives the surrounding world. Because she is the only agent, these possible worlds deal only with propositional facts about the surrounding world. Now, because we suppose that there are other agents than agent Y, a possible world for Y in that case should also deal with how the other agents perceive the surrounding world. These “multi-agent” possible worlds should then not only deal with propositional facts but also with epistemic facts. So to represent the other agents’ beliefs (possibly about agent Y’s beliefs) in a multi-agent possible world, we introduce a modal structure to our possible worlds. We do so as follows.

Definition 3.3 (Multi-agent possible world)
A multi-agent possible world (M, w) is a *finite* epistemic model $M = (W, \{R_j \mid j \in G\}, val)$ generated by w such that for all j, R_j is serial, transitive and euclidean, and

- $R_Y(w) = \{w\}$;

[2] Note that van Benthem, in [17], already mentioned that this proposition could be used in belief revision theory.

- there is no v and $j \neq Y$ such that $w \in R_j(v)$.

Let us have a closer look at the definition. The second condition will be motivated in the next section. The first condition ensures us that in case Y is the only agent then a multi-agent possible world boils down to an interpretation, as in the AGM theory. The first condition also ensures us that in case Y assumes that she is in the multi-agent possible world (M, w) then for her w is the only possible world. In fact the other possible worlds of a multi-agent possible world are just present for technical reasons: they express the other agents' beliefs (in world w). One could get rid of the condition that a multi-agent possible world (M, w) is generated by w but the worlds which do not belong to the submodel generated by w would not have neither philosophical nor technical motivation. Note that if we remove the constraints on the accessibility relations (seriality, euclidicity and transitivity) the results in this paper are still valid. We prefer to keep them because we find them more intuitive to model the notion of belief construed as conviction. Intuitively, this notion of belief corresponds for example to the kind of belief in a theorem that you have after having proved this theorem and checked the proof several times. In the literature, this notion of belief corresponds to Lenzen's notion of conviction [16], to Gärdenfors' notion of acceptance [10] and to Voorbraak's notion of rational introspective belief [19].

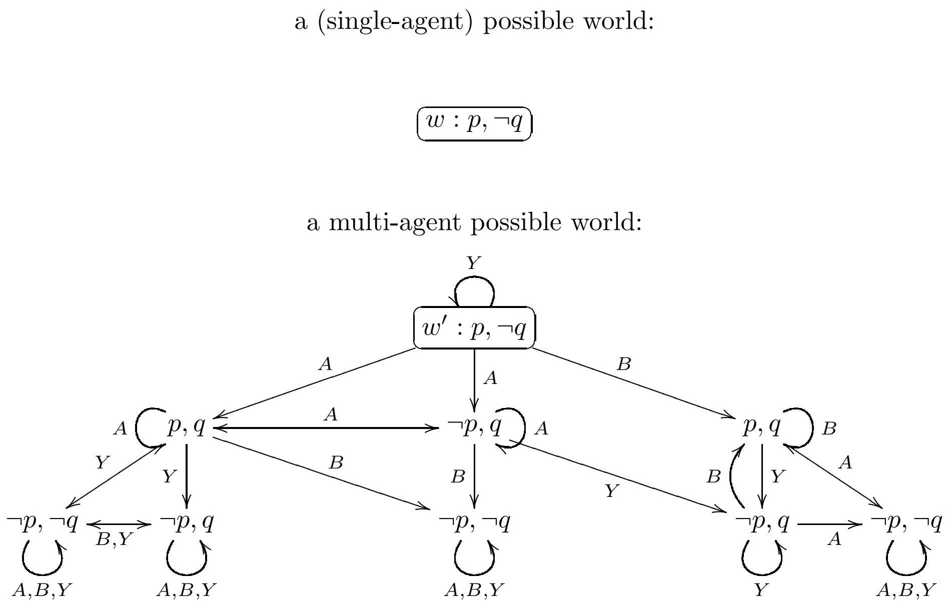

Fig. 2. From possible world to multi-agent possible world

We see in Figure 2 that a multi-agent possible world is really a generalization of a possible world (or interpretation). The epistemic state of the agent being represented semantically by a finite set of possible worlds (or interpretations) in AGM, this leads us to introduce the analogue notion of internal model in our multi-agent setting.

Definition 3.4 (Internal model)
An internal model is a finite and disjoint union of multi-agent possible worlds.

Note that in the single-agent case, an internal model boils down to a (non-empty) set of interpretations, so represents a belief set. Intuitively, an internal model is the formal model that agent Y has "in her head" and that represents how she perceives the surrounding world. This interpretation differs from Hintikka epistemic models (M, w_a), usually encountered in epistemic logic, which are supposed to represent objectively and from an external point of view how all the agents perceive the actual world w_a.

Example 3.5
An example of internal model is depicted in Figure 3. In this internal model, agent Y does not know wether p is true or not (formally $\neg B_Y p \wedge \neg B_Y \neg p$). Indeed, $M_1, w \models p$ and $M_2, v \models \neg p$. Agent Y also believes that the agent A does not know whether p is true or false (formally $B_Y(\neg B_A p \wedge \neg B_A \neg p)$). Indeed, $M_1, w \models \neg B_A p \wedge \neg B_A \neg p$ and $M_2, v \models \neg B_A p \wedge \neg B_A \neg p$. Finally, agent Y believes that A believes that she does not know whether p is true or false (formally $B_Y B_A(\neg B_Y p \wedge \neg B_Y \neg p)$) since $M_1, w \models B_A(\neg B_Y p \wedge \neg B_Y \neg p)$ and $M_2, v \models B_A(\neg B_Y p \wedge \neg B_Y \neg p)$.

3.2.2 Alternative representation of internal models

Definition 3.6
Let $\{(M_1, w_1), \ldots, (M_n, w_n)\}$ be an internal model, where $M_i = (W_i, \{R^i_j \mid j \in G\}, val_i)$. The *epistemic model associated to* $\{(M_1, w_1), \ldots, (M_n, w_n)\}$ is the $KD45_G$ epistemic model $M = (W, \{R_j \mid j \in G\}, val)$ defined as follows.

- $W = W_1 \cup \ldots \cup W_n$;
- $R_j = R^1_j \cup \ldots \cup R^n_j$ for $j \neq Y$;
- $R_Y = R^1_Y \cup \ldots \cup R^n_Y \cup \{(w_i, w_k); i, k = 1 \ldots n\}$;
- $val(w) = val_i(w)$ if $w \in W_i$.

Example 3.7
The internal model $\{(M_1, w), (M_2, v)\}$ is represented in Figure 3 and an epistemic model bisimilar to the epistemic model associated to $\{(M_1, w), (M_2, v)\}$ of Figure 4 is represented in Figure 5.

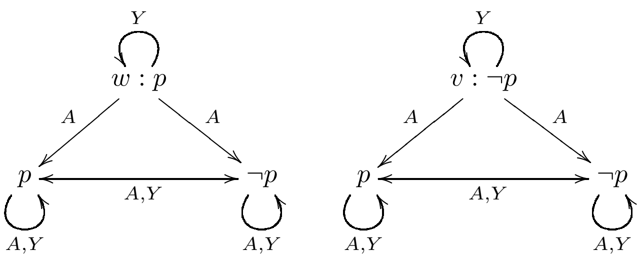

Fig. 3: An internal model : multi-agent possible world (M_1, w) (left) and multi-agent possible world (M_2, v) (right)

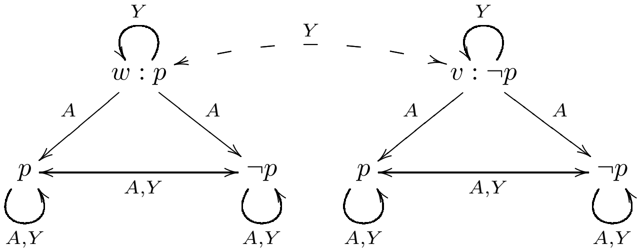

FIG. 4. Epistemic model associated to the internal model $\{(M_1, w), (M_2, v)\}$

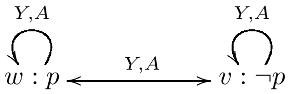

FIG. 5. Epistemic model bisimilar to the epistemic model of Figure 4

We can now motivate the second item of Definition 3.3. Indeed, if this item was not fulfilled then part of agent j's beliefs about Y's beliefs (for $j \neq Y$) would depend on the other multi-agent possible worlds of the internal model. This aspect of the notion of internal model is revealed when we define the notion of epistemic model associated to an internal model. Condition 2 ensures us that agents j's beliefs in a multi-agent possible world of a given internal model depend only on the structure of this multi-agent possible world. Condition 2 thus provides a kind of modularity to multi-agent possible worlds that will be useful in the sequel.

For every internal model, the epistemic model associated to this internal model is a $KD45_G$ epistemic model generated by the set of worlds $R_Y(w)$. The other way round, one can easily show that any $KD45_G$ epistemic model generated from $R_Y(w)$ (for some world w of this epistemic model) can be equivalently represented by an internal model.[3] So we have two equivalent ways to represent the epistemic state of agent Y.

The second type of representation is much closer to usual epistemic models of standard epistemic logic. But we stress that the interpretation of our models are different from the interpretation of epistemic models in standard epistemic logic. Our models are built by agent Y in order to represent for herself the surrounding world, whereas the models of epistemic logic are built by an external modeler and represent truthfully how *all* the agents perceive the actual world. Formally, the main difference is that they have a single actual world whereas in our internal models we have a set of 'actual worlds' (the roots of the multi-agent possible worlds) representing agent Y's uncertainty about the actual world.

Besides, the shape of internal models, based on the notion of multi-agent possible world, allows to generalize easily concepts and methods from AGM belief revision

[3]This equivalence could be easily specified formally by stating that for all i and $\phi \in \mathcal{L}^C_{\neq Y}$, $M, w_i \models \phi$ iff $M_i, w_i \models \phi$, where $\mathcal{L}^C_{\neq Y}$ is defined in Definition 3.8.

theory, as we will now see.

3.3 The multi-agent generalization of the AGM approach

In the multi-agent case like in the single-agent case, it does not make any sense to revise by formulas dealing with what agent Y believes or considers possible. Indeed, due to the fact that positive and negative introspection are valid in KD45, Y already knows all she believes and all she disbelieves. So we restrict the epistemic language to a fragment that we call $\mathcal{L}^C_{\neq Y}$ defined as follows.

Definition 3.8

$$\mathcal{L}^C_{\neq Y} : \phi := \top \mid p \mid B_j\psi \mid \phi \wedge \phi \mid \neg\phi,$$

where ψ ranges over $\mathcal{L}^C$ and j over $G - \{Y\}$.

Note that this definition does not rule out formulas of the form $B_A B_Y p$ which deal with agent A's beliefs *about* agent Y's beliefs.

We can then apply with some slight modifications the procedure spelled out for the single agent case in Section 2.

First the postulates for multi-agent belief revision are identical to the ones spelled out in Lemma 2.2 but this time ψ, ϕ and ϕ' belong to $\mathcal{L}^C_{\neq Y}$.

Now we define $\mathcal{I}_G$ to be the set of all multi-agent possible worlds modulo bisimulation, and we pick the smallest multi-agent possible world among each class of bisimilarly indistinguishable multi-agent possible worlds. We define the set of models associated to ψ by

$$Mod(\psi) = \{(M, w) \in \mathcal{I}_G \mid M, w \models \psi\}.$$

Let $\mathcal{M}$ be an internal model. Thanks to Proposition 3.2 we can easily prove the following fact.

Fact (*) There is a formula $form(\mathcal{M}) \in \mathcal{L}^C_{\neq Y}$ such that $Mod(form(\mathcal{M})) = \mathcal{M}$.

We then get the multi-agent generalization of Theorem 2.3 by replacing interpretations I by multi-agent possible worlds (M, w).

Theorem 3.9 (Representation theorem)
Revision operator $\circ$ on $\mathcal{L}^C_{\neq Y}$ satisfies conditions $(R1) - (R6)$ iff there exists a faithful assignment that maps each belief base ψ to a total pre-order $\leq_\psi$ such that $Mod(\psi \circ \phi) = Min(Mod(\phi), \leq_\psi)$.

Proof. The proof follows the line of that of Theorem 2.3. It relies heavily on the fact (*).

The "if" direction is straightforward. For the "only-if" direction, the key is the definition of a faithful assignment for each belief base in terms of $\circ$. For any multi-agent possible worlds (M, w) and (M', w') ($(M, w) = (M', w')$ is permitted), we define a relation $\leq_\psi$ as $(M, w) \leq_\psi (M', w')$ iff either $(M, w) \in Mod(\psi)$ or $(M, w) \in Mod(\psi \circ form(\{(M, w), (M', w')\})$.

This definition of the assignment is identical to the single agent case (see proof of Theorem 2.3). ∎

This similarity between Theorem 2.3 and Theorem 3.9 is depicted in Figure 6. We see in this figure that possible worlds of AGM belief revision are just replaced by multi-agent possible worlds which are represented by triangles.

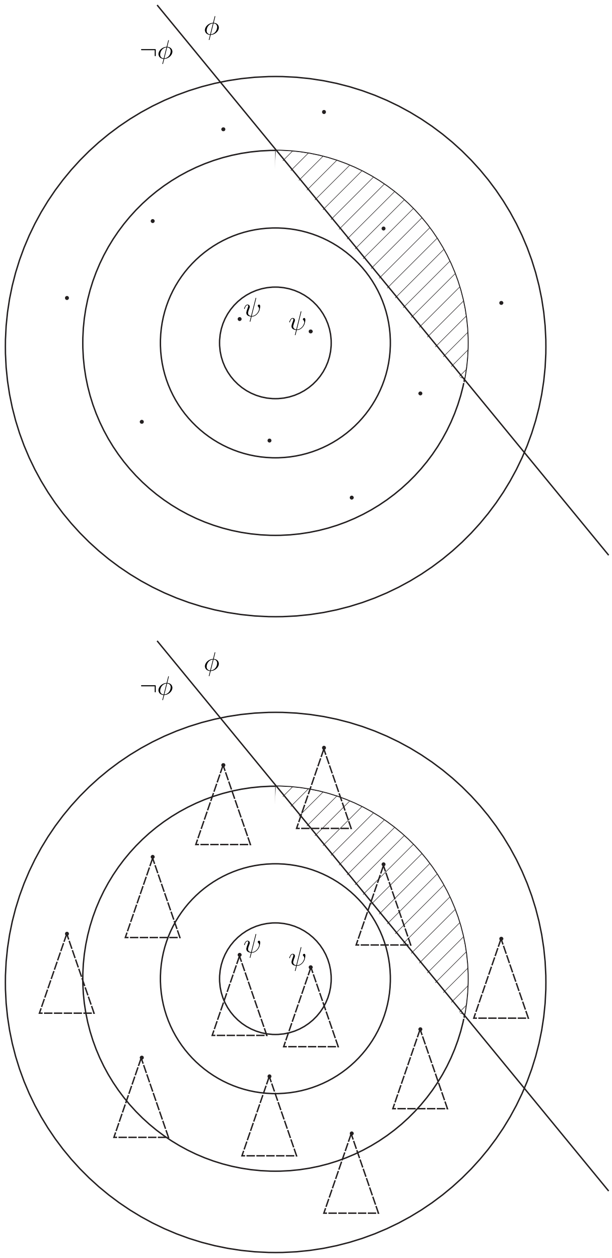

Fig. 6. AGM belief revision (*above*) and private multi-agent belief revision (*below*)

Remark 3.10 (Important)
We have picked only one of the theorems of [14] but in fact all the theorems present in [14] transfer to the multi-agent case. It includes in particular the following theorem, where $\leq_\psi$ is a partial order instead of a total order:

Revision operator $\circ$ satisfies postulates $(R1)-(R5)$, $(R7)$ and $(R8)$ if and only if there exists a faithful assignment that maps each belief base ψ to a *partial* pre-order $\leq_\psi$ such that $Mod(\psi \circ \phi) = Min(Mod(\phi), \leq_\psi)$; where

(R7) If $\psi \circ \phi_1 \rightarrow \phi_2$ and $\psi \circ \phi_2 \rightarrow \phi_1$ then $\psi \circ \phi_1 \leftrightarrow \psi \circ \phi_2$.
(R8) $(\psi \circ \phi_1) \wedge (\psi \circ \phi_2) \rightarrow \psi \circ (\phi_1 \vee \phi_2)$.

In summary, the concept of internal model allows for a straightforward transfer of the AGM framework and results.

4 Some considerations specific to our multi-agent approach

In this section we are going to investigate some multi-agent rationality postulates. Indeed, because we add a multi-agent structure to our possible worlds, it is natural to study how (agent Y's beliefs about) the other agents' beliefs evolve during a revision process.

As said in the introduction, the events we study are private announcements made to Y, the other agents not being aware of anything. So the beliefs of the other agents actually do not change and agent Y knows this. Consequently, agent Y's beliefs about the agents who are not concerned by the formula announced to her should not change as well. So, first of all, we need to define formally what are the agents who are concerned by a formula.

4.1 On the kind of information a formula is about

First note that an input may not only concern agents but also the objective state of nature, i.e. propositional facts, that we note pf. For example, the formula $p \wedge B_j B_i \neg p$ concerns agent j's beliefs but also propositional facts (namely p). Besides, a formula cannot be about Y's beliefs because $\phi \in \mathcal{L}^C_{\neq Y}$ by assumption. So what an input is about includes propositional facts but excludes agent Y's beliefs. This leads us to the following definition. Let $C = (G \cup \{\texttt{pf}\}) - \{Y\}$.

Definition 4.1
We define by induction the agents who are *concerned* by a formula as follows:

- $C(p) = \texttt{pf}$; $C(B_j\phi) = \{j\}$; $C(C_{G_1}\phi) = G_1$;
- $C(\neg\phi) = C(\phi)$; $C(\phi \wedge \phi') = C(\phi) \cup C(\phi')$.

For example, $C(p \vee (q \wedge B_j B_i r) \wedge B_k r)) = \{\texttt{pf}, j, k\}$, and $C(B_i p \vee B_j B_k \neg p) = \{i, j\}$.

We then define a language $\mathcal{L}^C_{C_1}$ whose formulas concern only agents in C_1, and possibly propositional facts if $\texttt{pf} \in C_1$.

Definition 4.2
Let $C_1 \subseteq C$. We define the language $\mathcal{L}^C_{C_1}$ as follows.

$$\phi = \top \mid p \mid B_j\psi \mid \phi \wedge \phi \mid \neg\phi,$$

where j ranges over C_1, ψ over formulas of $\mathcal{L}^C$ and p over A, where $A = \Phi$ if $\texttt{pf} \in C_1$ and $A = \emptyset$ otherwise.

Now we define a notion supposed to tell us whether two pointed and finite epistemic models contain the same information about some agents' beliefs and possibly about propositional facts.

DEFINITION 4.3
Let $C_1 \subseteq C$. We say that (M, w) and (M', w') are C_1-bisimilar, noted $M, w \leftrightarrows_{C_1} M', w'$, iff

- if $\texttt{pf} \in C_1$ then $val(w) = val(w')$ and
- for all $j_0 \in C_1$,
  if $v \in R_{j_0}(w)$ then there exists $v' \in R_{j_0}(w')$ such that $M, v \leftrightarrows M', v'$,
  if $v' \in R_{j_0}(w')$ then there exists $v \in R_{j_0}(w)$ such that $M, v \leftrightarrows M', v'$.

PROPOSITION 4.4
Let $C_1 \subseteq C$. Then $M, w \leftrightarrows_{C_1} M', w'$ iff for all $\phi \in \mathcal{L}^C_{C_1}$, $M, w \models \phi$ iff $M', w' \models \phi$.

Proposition 4.4 ensures us that the notion we just defined captures what we wanted. Its proof uses that the models are finite (otherwise the if direction would not hold). We then have a counterpart of Proposition 3.2.

PROPOSITION 4.5
Let $C_1 \subseteq C$, let M be a finite epistemic model and $w \in M$. Then there is $\delta^{C_1}_M(w) \in \mathcal{L}^C_{\neq Y}$ such that

1. $M, w \models \delta^{C_1}_M(w)$;
2. for every finite epistemic model M' and world $w' \in M'$, if $M', w' \models \delta^{C_1}_M$ then $M, w \leftrightarrows_{C_1} M', w'$.

DEFINITION 4.6
Let $\mathcal{M}$ and $\mathcal{M}'$ be two sets of multi-agent possible worlds. We set $\mathcal{M} \leftrightarrows_{C_1} \mathcal{M}'$ iff for all $(M, w) \in \mathcal{M}$ there exists $(M', w') \in \mathcal{M}'$ such that $M, w \leftrightarrows_{C_1} M', w'$, and for all $(M', w') \in \mathcal{M}'$ there exists $(M, w) \in \mathcal{M}$ such that $M, w \leftrightarrows_{C_1} M', w'$.

4.2 Some postulates specific to our multi-agent approach

As we said before, we study private announcement made to Y, the other agents not being aware of anything. So, in particular, Y's beliefs about the beliefs of the agents who are not concerned by the formula should not change. This can be captured by the following postulate:

(RG1) Let $\phi, \phi' \in \mathcal{L}^C_{\neq Y}$ such that $C(\phi) \cap C(\phi') = \emptyset$.
If $\psi \to \phi'$ then $(\psi \circ \phi) \to \phi'$

This postulate is the multi-agent version of Parikh and Chopra's postulate [8]. The example of the introduction illustrates this postulate: there $\phi = \neg p$ and $\phi' = B_j p \wedge B_j C_G p$. Now the semantic counterpart of $(RG1)$:

PROPOSITION 4.7
Revision operator $\circ$ satisfies $(RG1)$ iff for all $\phi \in \mathcal{L}^C_{\neq Y}$, for all $(M', w') \in Mod(\psi \circ \phi)$ there is $(M, w) \in Mod(\psi)$ such that $M, w \leftrightarrows_{C'} M', w'$, with $C' = C - C(\phi)$.

Let us consider the converse of ($RG1$).

(RG2) Let $\phi, \phi' \in \mathcal{L}^C_{\neq Y}$ such that $C(\phi) \cap C(\phi') = \emptyset$.
If $\psi \wedge \phi'$ is satisfiable then $(\psi \circ \phi) \wedge \phi'$ is satisfiable.

And the semantic counterpart:

PROPOSITION 4.8
Revision operator $\circ$ satisfies ($RG2$) iff for all $\phi \in \mathcal{L}^C_{\neq Y}$, for all $(M, w) \in Mod(\psi)$ there is $(M', w') \in Mod(\psi \circ \phi)$ such that $M, w \leftrightarrows_{C'} M', w'$, with $C' = C - C(\phi)$.

Unlike ($RG1$), ($RG2$) is not really suitable for revision because all the worlds representing Y's epistemic state "survive" the revision process if ($RG2$) is fulfilled. This should not be the case in general because new information can discard some previous possibilities. This is however the case for update where we apply the update process to each world independently (see [13] for an in depth analysis). So ($RG2$) is more suitable for an update operator.

In fact ($RG2$) is similar to the propositional update postulate $(U8)(\psi \vee \psi') \circ \phi \leftrightarrow (\psi \circ \phi) \vee (\psi' \circ \phi')$. For example, consider $\psi = B_i p \vee B_j p$ and $\phi = \neg B_i p$. Then the revised formula is $\psi \circ \phi = B_j p \wedge \neg B_i p$ according to postulate ($R2$). But according to postulate ($RG2$), $\neg B_j p$ should be satisfiable after the revision because $\psi \wedge \neg B_j p$ was satisfiable before.

Postulates ($RG1$) and ($RG2$) together are equivalent to: for all $\phi, \phi' \in \mathcal{L}^C_{\neq Y}$ such that $C(\phi) \cap C(\phi') = \emptyset$, $\psi \to \phi'$ iff $(\psi \circ \phi) \to \phi'$. Then

PROPOSITION 4.9
Revision operator $\circ$ satisfies ($RG1$) and ($RG2$) iff for all $\phi \in \mathcal{L}^C_{\neq Y}$, $Mod(\psi) \leftrightarrows_{C'} Mod(\psi \circ \phi)$, with $C' = C - C(\phi)$.

5 An example of revision operator

In this section we propose a revision operator based on a degree of similarity between multi-agent possible worlds defined very much in the same way as in [15]. Besides, for sake of simplicity, we assume that formulas representing belief bases and private announcements belong to the language associated to Y *without* common belief, noted $\mathcal{L}_{\neq Y}$:

$$\mathcal{L}_{\neq Y} : \phi := \top \mid p \mid B_j \psi \mid \phi \wedge \phi \mid \neg \phi,$$

where ψ ranges over $\mathcal{L}$ and j over $G - \{Y\}$. One should note that in this setting, the "if" direction of Theorem 3.9 still holds, but not the "only if" direction.

5.1 Mathematical preliminaries

5.1.1 Lexicographic ordering

We first recall the definition of an anti-lexicographic ordering.

DEFINITION 5.1
Let $k \in \mathbb{N}$ and $(l_0, \ldots, l_k), (l'_0, \ldots, l'_k) \in [0; 1]^{k+1}$. We set

$$(l_0,\dots,l_k) < (l'_0,\dots,l'_k) \quad \text{iff} \quad \begin{cases} l_k < l'_k \text{ or} \\ l_k = l'_k,\dots,l_{k-j+1} = l'_{k-j+1} \text{ and} \\ l_{k-j} < l'_{k-j} \text{ for some } 1 \leq j \leq k. \end{cases}$$

Now we define the *Supremum* of a set of tuples with respect to the anti-lexicographic ordering by using the supremum *Sup* of real numbers.

Definition 5.2
Let $k \in \mathbb{N}$ and $\{(l^i_0,\dots,l^i_k) \mid i \in S\} \subseteq [0;1]^{k+1}$ (where S is an index set which is possibly infinite). $Sup^k\{(l^i_0,\dots,l^i_k) \mid i \in S\} = (A_0,\dots,A_k)$ is defined as follows.

$$A_k = Sup\{l^i_k \mid i \in S\}; \text{and for all } m < k,$$

$$A_m = \begin{cases} Sup\{l^i_m \mid l^i_j = A_j \text{ for all } k \geq j > m, i \in S\} \\ \text{if there is } i \text{ such that } l^i_j = A_j \text{ for all } k \geq j > m \\ Sup\{l^i_m \mid i \in S\} \\ \text{otherwise.} \end{cases}$$

where *Sup* is the usual supremum on real numbers.

This definition is well-founded because the supremum of a non-empty set of real numbers with an upper bound always exists. Finally, we check that this supremum of tuples does correspond to the maximum of tuples when this one exists.

Proposition 5.3
Let $L = \{(l^i_0,\dots,l^i_k) \mid i \in S\} \subseteq [0;1]^{k+1}$ and $(l^{i_0}_0,\dots,l^{i_0}_k) \in L$ (where S is an index set which is possibly infinite).

$$\text{If } (l^{i_0}_0,\dots,l^{i_0}_k) \geq (l^i_0,\dots,l^i_k) \text{ for all } i \in S, \text{ then } (l^{i_0}_0,\dots,l^{i_0}_k) = Sup^k(L).$$

5.1.2 n-bisimulation

Our definition of n-bisimulation is a slight modification of the definition of n-bisimulation in [4] [7].

Definition 5.4
Let $M = (W, R, val)$ and $M' = (W', R', val')$ be two epistemic models, and let $w \in M$, $w' \in M'$. Let $Z \subseteq W \times W'$. We recursively define the property of Z being n-bisimulation in w and w', noted $Z : M, w \leftrightarrows_n M', w'$:

1. $Z : M, w \leftrightarrows_0 M', w'$ iff wZw' and $val(w) \neq val(w')$;
2. $Z : M, w \leftrightarrows_1 M', w'$ iff wZw' and $val(w) = val(w')$;
3. For all $n \geq 1$, $Z : M, w \leftrightarrows_{n+1} M', w'$ iff wZw' and $val(w) = val(w')$ and for all $j \in G$,
   - for all $v \in R_j(w)$ there is $v' \in R_j(w')$ such that $Z : M, v \leftrightarrows_n M', v'$.
   - for all $v' \in R_j(w')$ there is $v \in R_j(w)$ such that $Z : M, v \leftrightarrows_n M', v'$.

Now we can define n-bisimilarity between w and w', noted $M, w \leftrightarrows_n M', w'$ by $M, w \leftrightarrows_n M', w'$ iff there exists a relation Z such that $Z : M, w \leftrightarrows_n M', w'$.

Two worlds being n-bisimilar (with $n \geq 1$) intuitively means that they have the same modal structure up to modal depth $n-1$, and thus they satisfy the same formulas of degree at most $n-1$. For example, in the epistemic models of Figure 7, we have $M, w \underline{\leftrightarrow}_1 M', w'$, but $M, w \underline{\leftrightarrow}_2 M', w'$ is not the case.

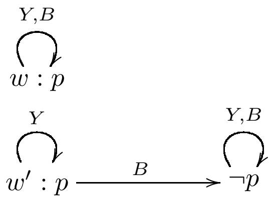

FIG. 7. Epistemic model (M, w) (*above*) and (M', w') (*below*)

The usual definition of Z being a bisimulation corresponds to $Z : M, w \underline{\leftrightarrow}_n M', w'$ for all $n \in \mathbb{N}^*$ ($\mathbb{N}^* = \mathbb{N} - \{0\}$). In fact, it suffices that two finite epistemic models be n-bisimilar up to a certain modal depth to be bisimilar, as the following proposition shows.

PROPOSITION 5.5
[3] Let M and M' be two finite epistemic models and $w \in M$, $w' \in M'$. Let $n = |M| \cdot |M'| + 1$. Then,

$$M, w \underline{\leftrightarrow}_n M', w' \text{ iff } M, w \underline{\leftrightarrow} M', w'.$$

5.2 Definition of the revision operator

First we are going to define a degree of similarity between two multi-agent possible worlds that will allow for an anti-lexicographic order.

DEFINITION 5.6
Let (M, w) and (M', w') be two multi-agent possible worlds, let $v \in M$ and $v' \in M'$, let S and S' be two finite sets of possible worlds, and let $\mathcal{M}$ and $\mathcal{M}'$ be two sets of multi-agent possible worlds (possibly infinite). Let $n = |M| \cdot |M'| + 1$ and $k \in \mathbb{N}$.

If E is a finite set of real numbers, we note $m(E)$ the average of E, i.e. $m(E) = \frac{1}{|E|} \sum_{e \in E} e$.

- $\sigma(v, v') = max\{\frac{i}{n} \mid M, v \underline{\leftrightarrow}_i M', v' \text{ and } i \in \{0, \ldots, n\}\}$;
- $\sigma(S, S') = \frac{1}{2}\left(m\{\sigma(s, S') \mid s \in S\} + m\{\sigma(S, s') \mid s' \in S'\}\right)$
  where $\sigma(s, S') = max\{\sigma(s, s') \mid s' \in S'\}$ and $\sigma(S, s') = max\{\sigma(s, s') \mid s \in S\}$;
- $s^k((M, w), (M', w')) = (\sigma(w, w'), m\{\sigma(R_j(w), R_j(w')) \mid j \in G, j \neq Y\}, \ldots,$
  $m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_k}(w), R_{j_1} \circ \ldots \circ R_{j_k}(w')) \mid j_1, \ldots, j_k \in G, j_i \neq j_{i+1}, j_1 \neq Y\})$;
- $s^k(\mathcal{M}, \mathcal{M}') = Sup^k\{s^k((M, w), (M', w')) \mid (M, w) \in \mathcal{M}, (M', w') \in \mathcal{M}'\}$.

$\sigma(v, v')$ measures a degree of similarity between the worlds v and v'. For example in Figure 7, we have $\sigma(w, w') = \frac{1}{3}$. Note that $0 \leq \sigma(v, v') \leq 1$ for all v and v'. If $\sigma(v, v') = 1$ then the worlds v and v' are bisimilar by Proposition 5.5. So their degree of similarity is the highest possible. If $\sigma(v, v') = 0$, that is $M, v \underline{\leftrightarrow}_0 M', v'$ then their degree of similarity is the lowest possible because they differ even on propositional facts. Likewise, $\sigma(S, S')$ measures a degree of similarity between the sets of worlds S and S'. Note also that $0 \leq \sigma(S, S') \leq 1$ for all S and S'. If $\sigma(S, S') = 1$ then for all worlds $v \in S$ there is $v' \in S'$ such that v is bisimilar with v', and vice versa, for all $v' \in S'$ there is $v \in S$ such that v' is bisimilar with v. So the degree of similarity between S and S' is the highest possible. If $\sigma(S, S') = 0$ then for all $v \in S$ there is no $v' \in S'$ such that v and v' agree on all propositional letters, and vice versa, for all $v' \in S'$ there is no $v \in S$ such that v and v' agree on all propositional letters. So the degree of similarity is the lowest possible. To be more precise, $\sigma(v, S')$ is the degree of similarity of a world v with S'. So $m\{\sigma(v, S') \mid v \in S\}$ is the average degree of similarity of a world $v \in S$ with S'. Likewise, $m\{\sigma(S, v') \mid v' \in S'\}$ is the average degree of similarity of a world $v' \in S'$ with S. So the degree of similarity between S and S' is just the average of these two degrees. $s^k((M, w), (M', w'))$ is a tuple which represents by how much two multi-agent possible worlds are similar relatively to their respective modal depth. For example in Figure 7 we have $s^2((M, w), (M', w')) = (\frac{1}{3}, 0, 0)$. Note that for a given modal depth we only compare the degree of similarity of worlds which have the same history (i.e. they are all accessed from w and w' by the same sequence of accessibility relations $R_{j_1}, \ldots, R_{j_k}$). Doing so, in our comparison we stick very much to the modal structure of both multi-agent possible worlds. We also assume that $j_i \neq j_{i+1}$ because otherwise, by transitivity and euclidicity of the accessibility relations, we would have $R_{j_i} = R_{j_i} \circ R_{j_{i+1}}$. Besides we take the average of their degree of similarity for every possible history in order to give the same importance to these different possible histories.

Definition 5.7
Let $\psi \in \mathcal{L}_{\neq Y}$ and $k = deg(\psi) + 1$. We assign to ψ a total pre-order $\leq_\psi$ on multi-agent possible worlds defined as follows:

$$(M, w) \leq_\psi (M', w') \text{ iff } s^k(Mod(\psi), \{(M, w)\}) \geq s^k(Mod(\psi), \{(M', w')\}).$$

The revision operator $\circ$ associated to this pre-order $\leq_\psi$ is defined semantically in the usual way (see Theorem 2.3) by:

$$Mod(\psi \circ \phi) = Min(Mod(\phi), \leq_\psi).$$

So (M, w) is closer to ψ than (M', w') when its degree of similarity with the models of ψ is higher than the degree of similarity of (M', w') with the models of ψ. In the next section, we are going to motivate our use of anti-lexicographic ordering and explain why we compare the modal structures of the multi-agent possible worlds only until modal depth $k = deg(\psi) + 1$.

5.3 Properties of the revision operator

Proposition 5.8
Let (M, w) be a multi-agent possible world and $\psi \in \mathcal{L}_{\neq Y}$ a satisfiable formula such that $deg(\psi) = d$. Then there is $(M_\psi, w_\psi) \in Mod(\psi)$ such that

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w), R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w_\psi)) \mid j_1, \ldots, j_{d+1} \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1.$$

This proposition tells us that, given a formula ψ of degree d and a multi-agent possible world (M, w), there is a multi-agent possible world that satisfies ψ and whose structure is the same as (M, w) beyond modal depth d. That is why, in $s^k(Mod(\psi), (M, w))$, we stop at modal depth $k = d + 1$ when we compare models of ψ with (M, w): we know that there is anyway a model of ψ whose modal structure is the same as (M, w) beyond this modal depth, so there is no need to check it further. Moreover, we would like to give priority to this similarity when we compare models of ψ with (M, w). That is to say, we would like to ensure that the models of ψ closest to (M, w) are such that their modal structure beyond this modal depth is the same as the one of (M, w). We do so by using the anti-lexicographic order defined in Definition 5.1.

The following proposition shows that we need to consider only finitely many models of ϕ in $s^k(Mod(\psi), (M, w)) = Sup^k\{s^k((M', w'), (M, w)) \mid (M', w') \in Mod(\phi)\}$.

Proposition 5.9
Let (M, w) be a multi-agent possible world. For all $k \in \mathbb{N}^*$, there are finitely many multi-agent possible worlds (M', w') such that

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_k}(w'), R_{j_1} \circ \ldots \circ R_{j_k}(w)) \mid j_1, \ldots, j_k \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1.$$

Corollary 5.10
Let (M, w) be a multi-agent possible world, $\psi \in \mathcal{L}_{\neq Y}$ and $k = deg(\psi) + 1$. Then there is $(M', w') \in Mod(\psi)$ such that $s^k((M', w'), (M, w)) = s^k(Mod(\psi), \{(M, w)\})$.

In other words, this corollary tells us that $s^k(Mod(\psi), \{(M, w)\}) = Sup^k\{s^k((M', w'), (M, w)) \mid (M', w') \in Mod(\psi)\}$ is actually a maximum.

Finally, we have the following nice property.

Proposition 5.11
The assignment defined in Definition 5.7 is a faithful assignment. Therefore the operator $\circ$ defined in Definition 5.7 satisfies the postulates $(R1) - (R6)$. Besides, $\circ$ satisfies also postulates $(RG1)$.

Note that postulate $(RG2)$ is not necessarily satisfied. This is reassuring because, as we said after Proposition 4.8, postulate $(RG2)$ is not really suitable for revision operators.

5.4 Concrete example

The revision operators $\circ$ we introduced so far were syntactic. But in fact we could also define revision operators directly on internal models. Indeed, as we said internal models are formal representations that agent Y has 'in her mind'. So we need revision mechanisms that she could use to revise her formal representation when she receives an input under the form of an epistemic formula. Such revision operators would then take an internal model and an input formula as arguments and would yield another internal model. The following definition gives an example of such a revision operator.

Definition 5.12
Let $\mathcal{M}$ be a set of multi-agent possible worlds and $\phi \in \mathcal{L}_{\neq Y}$ a satisfiable formula. We define the revision of $\mathcal{M}$ by ϕ, noted $\mathcal{M} * \phi$, as follows.

$$\mathcal{M} * \phi = Min\,(Mod(\phi), \leq_{\mathcal{M}})$$

where for all multi-agent possible worlds (M, w) and (M', w'),

$$(M, w) \leq_{\mathcal{M}} (M', w') \text{ iff } s^k(\mathcal{M}, (M, w)) \geq^k s^k(\mathcal{M}, (M', w'))$$

where $k = deg(\phi) + 1$.

The reason why we stop at modal depth $k = deg(\phi) + 1$ is the same reason why we stopped at modal depth $k = deg(\psi) + 1$ for $s^k(Mod(\psi), (M, w))$ in Definition 5.7. It is because we know thanks to Proposition 5.8 that there is a model of ϕ and a multi-agent possible world of $\mathcal{M}$ which agree on their modal structure beyond modal depth $deg(\phi)$.

However, note that if $\mathcal{M}$ is an internal model then $\mathcal{M} * \phi$ might be infinite and therefore not an internal model. The following proposition ensures us that it is not the case.

Proposition 5.13
Let $\mathcal{M}$ be an internal model and $\phi \in \mathcal{L}_{\neq Y}$ a satisfiable formula. Then $\mathcal{M} * \phi$ is an internal model.

Example 5.14
Let us take up Example 3.5. Agent Y's initial internal model $\{(M, w), (M', w')\}$ is depicted in Figure 8. The epistemic model associated to $\{(M, w), (M', w')\}$ is depicted in Figure 9.

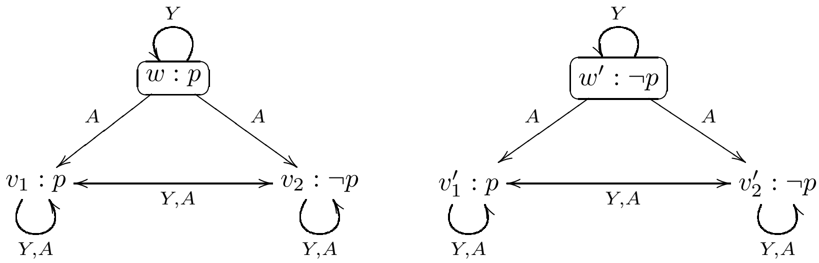

Fig. 8. Agent Y's initial internal model $\{(M, w), (M', w')\}$

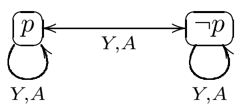

Fig. 9. Epistemic model associated to $\{(M, w), (M', w')\}$

Now, suppose that an external agent announces to agent Y *privately* that agent A believes p is true (formally $B_A p$). This announcement contradicts of course her beliefs and she has to revise her internal model. The following proposition tells us that the revised model is $\{(M^r, w^r), (M^{r'}, w^{r'})\}$, which is depicted in Figure 10.

Proposition 5.15
$\{(M,w),(M',w')\} * B_A p = \{(M^r,w^r),(M^{r'},w^{r'})\}$

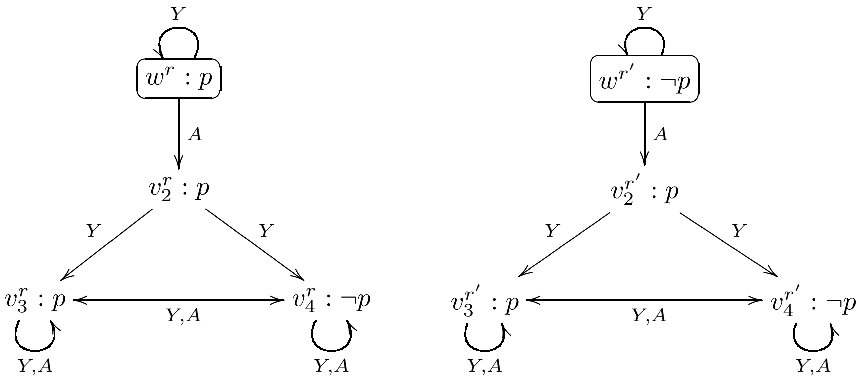

Fig. 10: Revised internal model $\{(M^r,w^r),(M^{r'},w^{r'})\}$ after the private announcement made to agent Y that agent A believes that p is true ($B_A p$)

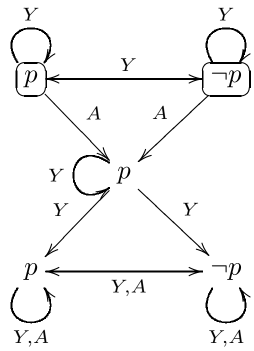

Fig. 11: Epistemic model bisimilar to the epistemic model associated to $\{(M^r,w^r),(M^{r'},w^{r'})\}$

The epistemic model associated to $\{(M^r,w^r),(M^{r'},w^{r'})\}$ is depicted in Figure 11. If we compare this internal model with the original internal model of Figure 9, we observe that agent Y still does not know whether p is true or not. This is what we should expect since the announcement was only about A's beliefs and did not give any information about the actual state of the world (as it would have been the case if the private announcement was that A *knows* that p is true). Of course, Y's beliefs about A's beliefs have changed because she now believes that A believes that the coin is heads up, unlike before. But (Y's beliefs about) A's beliefs about Y's beliefs have not changed. This is also what we should expect. Indeed, A is not aware of this private announcement to Y, so his beliefs about Y's beliefs do not change, and Y knows this. And because these beliefs are independent from his beliefs about propositional facts like p, Y's beliefs about A's beliefs about Y's beliefs should not change during the revision process. More generally, Y's beliefs about beliefs of degree larger than 1, i.e. larger than the degree of $B_A p$, should not change. Formally, this is exactly what our anti-lexicographic ordering and Proposition 5.8 ensure.

Remark 5.16
This example suggests that we could strengthen and refine our postulate ($RG1$) and require more demanding and more precise conditions. For example, if $\phi = p \wedge B_j B_i q \wedge B_i B_j B_i p$, then this formula is certainly about propositional facts, about agent j's beliefs, and about agent i's beliefs: $C(\phi) = \{\texttt{pf}, j, i\}$. But it is more precisely about propositional facts, about agent j's beliefs *about* agent i's beliefs, and about agent i's beliefs *about* agent j's beliefs about agent i's beliefs: $S(\phi) = \{\texttt{pf}, (j,i), (i,j,i)\}$. So, what should not change during a revision by ϕ are all beliefs ϕ' whose corresponding set of sequences $S(\phi')$ does not intersect with $S(\phi)$, which includes here all formulas of degree higher than 3 (because $deg(\phi) = 3$). Formally, this corresponds to refining ($RG1$) by the following postulate.

(RG1') Let $\phi, \phi' \in \mathcal{L}^C_{\neq Y}$ such that $S(\phi) \cap S(\phi') = \emptyset$.
If $\psi \rightarrow \phi'$ then $\psi \circ \phi \rightarrow \phi'$.

6 Conclusion

We have proposed a semantics to adequately represent agent Y's perception of the surrounding world in a multi-agent setting. This semantics generalizes the single agent one of AGM belief revision theory. Then Proposition 3.2 has enabled us to generalize easily the (representational) results of AGM belief revision theory to the multi-agent case. Finally, we have studied two additional multi-agent postulates and we have given an example of a concrete revision operator that satisfies one of these multi-agent postulates.

The power of our approach is that it generalizes the (representational) results of AGM belief revision theory to the multi-agent case, and so thanks to the notion of internal model. In fact, if we consider in particular that there are no other agents than Y then our approach boils down to classical AGM belief revision theory.

In the literature of dynamic epistemic logic, there are works that also deal with private multi-agent belief revision ([2],[5] or [18] for example). However, their modeling approach is quite different from ours. The models built in their work are supposed to represent truthfully the situation from an external and objective point of view, as it is usually done in epistemic logic. So the shape of their models is different from our internal models as we said in Section 3.2.2. In that respect, they also often introduce in their models possible worlds that the agents do not consider *consciously* as being possible but that are nevertheless relevant to model the agents' epistemic states from an external point of view (these possible worlds somehow express what would surprise the agents). The revision mechanisms they propose rely heavily on the existence of these possible worlds without whom revision would not be possible. On the other hand, we do not explicitly introduce these possible worlds in our formalism in order to perform belief revision. Indeed, as we said, the agents do not consider these worlds *consciously* as being possible whereas in our approach we intend to model (by means of internal models) only the beliefs of a particular agent Y, which are assumed to be conscious. This explains and motivates why our methods are different from theirs. However we do not claim that they are superior: they are simply different because our modeling approach is different.

Finally, it would be interesting to investigate other multi-agent postulates and other distances over multi-agent possible worlds. Another line of research would be to study multi-agent update as we have started in Section 4.2. Indeed, the results of [14] about propositional update transfer to the multi-agent case as well.

Acknowledgements

I thank my PhD supervisors Hans van Ditmarsch and Andreas Herzig for useful discussions and comments. I also thank Jérôme Lang for comments on an earlier version of this paper.

References

[1] Carlos Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985.

[2] Guillaume Aucher. A combined system for update logic and belief revision. In Mike Barley and Nikola K. Kasabov, editors, *PRIMA 2004*, volume 3371 of *LNCS*, pages 1–17. Springer, 2004. Revised Selected Papers.

[3] Philippe Balbiani. Manuscript, unpublished. 2007.

[4] Philippe Balbiani and Andreas Herzig. Talkin'bout Kripke models. In Torben Braüner and Jrgen Villadsen, editors, *International Workshop on Hybrid Logic 2007 (Hylo 2007)*, Dublin, 06/08/07-10/08/07 2007.

[5] Alexandru Baltag and Sonja Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electr. Notes Theor. Comput. Sci.*, 165:5–21, 2006.

[6] John Barwise and Larry Moss. *Vicious Circles*. CSLI Publications, Stanford, 1997.

[7] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Computer Science*. Cambridge University Press, 2001.

[8] Samir Chopra and Rohit Parikh. An inconsistency tolerant model for belief representation and belief revision. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 192–199, 1999.

[9] Ronald Fagin, Joseph Halpern, Yoram Moses, and Moshe Vardi. *Reasoning about knowledge*. MIT Press, 1995.

[10] Peter Gärdenfors. *Knowledge in Flux (Modeling the Dynamics of Epistemic States)*. Bradford/MIT Press, Cambridge, Massachusetts, 1988.

[11] Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[12] Jaakko Hintikka. *Knowledge and Belief, An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca and London, 1962.

[13] Hirofumi Katsuno and Alberto Mendelzon. On the difference between updating a knowledge base and revising it. In *Proceedings of Knowledge Representation*, pages 387–394, 1991.

[14] Hirofumi Katsuno and Alberto Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52(3):263–294, 1992.

[15] Daniel Lehmann, Menachem Magidor, and Karl Schlechta. Distance semantics for belief revision. *Journal of Symbolic Logic*, 66(1):295–317, 2001.

[16] Wolfgang Lenzen. *Recent Work in Epistemic Logic*. Acta Philosophica 30. North Holland Publishing Company, 1978.

[17] Johan van Benthem. "One is a Lonely Number": logic and communication. In Z. Chatzidakis, P. Koepke, and W. Pohlers, editors, *Logic Colloquium'02*, volume 27 of *Lecture Notes in Logic*. 2006. Association for Symboloc Logic.

[18] Hans van Ditmarsch. Prolegomena to dynamic logic for belief revision. *Synthese*, 147:229–275, 2005.

[19] Frans Voorbraak. *As Far as I know. Epistemic Logic and Uncertainty.* PhD thesis, Utrecht University, 1993.

A Proofs of Proposition 3.2 and Fact (*)

PROPOSITION A.1 (Proposition 3.2)
Let M be a *finite* epistemic model and $w \in M$. Then there is an epistemic formula $\delta_M(w)$ (involving common knowledge) such that

1. $M, w \models \delta_M(w)$
2. For all epistemic model M', w', if $M', w' \models \delta_M(w)$ then $M, w \leftrightarrows M', w'$.

PROOF. The proof can be found in [4]. ∎

Fact (*) there is a formula $form(\mathcal{M}) \in \mathcal{L}^C_{\neq Y}$ such that $Mod(form(\mathcal{M})) = \mathcal{M}$.

PROOF. Let (M, w) be a multi-agent possible world. Then we set

$$\delta^*_M(w) = \bigwedge_{p \in val(w)} p \wedge \bigwedge_{p \notin val(w)} \neg p \wedge \bigwedge_{j \in G - \{Y\}} \left(\bigwedge_{v \in R_j(w)} \neg B_j \neg \delta_M(v) \wedge B_j \left(\bigvee_{v \in R_j(w)} \delta_M(v) \right) \right).$$

Clearly $\delta^*_M(w) \in \mathcal{L}^C_{\neq Y}$, $M, w \models \delta^*_M(w)$ and for all multi-agent possible worlds (M', w'), if $M', w' \models \delta^*_M(w)$ then $M, w \leftrightarrows M', w'$ by applying Proposition 3.2. Let $\mathcal{M} = \{(M_1, w_1); \ldots; (M_n, w_n)\}$. We set $form(\mathcal{M}) = \delta^*_{M_1}(w_1) \vee \ldots \vee \delta^*_{M_n}(w_n)$. Then $form(\mathcal{M}) \in \mathcal{L}^C_{\neq Y}$ and $Mod(form(\mathcal{M})) = \mathcal{M}$ because $\mathcal{I}_G$ consists of all the multi-agent possible worlds *modulo bisimulation.* ∎

B Proofs of Propositions 4.4, 4.5, 4.7, 4.8 and 4.9

PROPOSITION B.1 (Proposition 4.4)

Let $C_1 \subseteq C$ $M, w \leftrightarrows_{C_1} M', w'$ iff for all $\phi \in \mathcal{L}^C_{C_1}$, $M, w \models \phi$ iff $M', w' \models \phi$.

PROOF. We assume that `pf` $\in C_1$, the proof without this assumption is essentially the same.

- Assume $M, w \leftrightarrows_{C_1} M', w'$. We are going to prove by induction on $\phi \in \mathcal{L}^C_{C_1}$ that $M, w \models \phi$ iff $M', w' \models \phi$.
  - $\phi = p$. As `pf` $\in C_1$, $M, w \models p$ iff $M', w' \models p$.
  - $\phi = \phi_1 \wedge \phi_2, \phi = \neg \phi'$ work by induction hypothesis.
  - $\phi = B_{j_1}\phi', j_1 \in C_1$. Assume $M, w \models B_{j_1}\phi'$ then for all $v \in R_{j_1}$, $M, v \models \phi'$ (*). But for all $v' \in R_{j_1}(w')$ there is $v \in R_{j_1}(w)$ such that $M, v \leftrightarrows M', v'$.
    So for all $v' \in R_{j_1}(w'), M', v' \models \phi'$ by property of the bisimulation and (*). Finally $M', w' \models B_{j_1}\phi'$, i.e. $M', w' \models \phi$.
    The other way around we can show that if $M', w' \models B_{j_1}\phi$ then $M, w \models B_{j_1}\phi$.
- Assume that for all $\phi \in \mathcal{L}^C_{C_1}$, $M, w \models \phi$ iff $M', w' \models \phi$ (*).
  - Clearly for all $p \in \Phi$, $w \in V(p)$ iff $w' \in V'(p)$.
  - Let $j_1 \in C_1$ and $v \in R_{j_1}(w)$.
    Assume for all $v' \in R_{j_1}(w')$ it is not the case that $M, v \leftrightarrows M', v'$ (**).
    Then for all $v' \in R_{j_1}(w')$ there is $\phi(v') \in \mathcal{L}$ such that $M, v \models \neg\phi(v')$ and $M', v' \models \phi(v')$.
    As by hypothesis W' is finite, let $\phi(w') = B_{j_1}\left(\bigvee_{v' \in R_{j_1}(w')} \phi(v') \right)$; then $\phi(w') \in \mathcal{L}^C_{C_1}$.
    Besides $M', w' \models \phi(w')$ but $M, w \models \neg\phi(w')$. This is impossible by (*), so (**) is false.
    The other part of the definition of $\leftrightarrows_{C_1}$ is proved similarly.

∎

Proposition B.2 (Proposition 4.5)
For $C_1 \subseteq C$, all pointed and finite epistemic model (M, w), there is $\delta_M^{C_1}(w)$ such that
- $M, w \models \delta_M^{C_1}(w)$
- for all pointed and finite epistemic model (M', w'), if $M', w' \models \delta_M^{C_1}$ then $M, w \leftrightarrows_{C_1} M', w'$

Proof. We only sketch the proof. If $\texttt{pf} \in C_1$, take

$$\delta_M^{C_1}(w) = \bigwedge_{\{p \mid w \in V(p)\}} p \wedge \bigwedge_{\{p \mid w \notin V(p)\}} \neg p \wedge \bigwedge_{j \in C_1} \left(\bigwedge_{v \in R_j(w)} \neg B_j \neg \delta_M(v) \wedge B_j \left(\bigvee_{v \in R_j(w)} \delta_M(v) \right) \right)$$

otherwise if $\texttt{pf} \notin C_1$, take

$$\delta_M^{C_1}(w) = \bigwedge_{j \in C_1} \left(\bigwedge_{v \in R_j(w)} \neg B_j \neg \delta_M(v) \wedge B_j \left(\bigvee_{v \in R_j(w)} \delta_M(v) \right) \right)$$ ∎

Proposition B.3 (Proposition 4.7)
Revision operator $\circ$ satisfies $(RG1)$ iff for all $\phi \in \mathcal{L}_{\neq Y}^{C}$, for all $(M', w') \in Mod(\psi \circ \phi)$ there is $(M, w) \in Mod(\psi)$ such that $M, w \leftrightarrows_{C'} M', w'$, with $C' = C - C(\phi)$.

Proof. The "if" part is straightforward. Let us prove the "only if" part. Let $\phi \in \mathcal{L}_{\neq Y}^{C}$ and let $(M', w') \in Mod(\psi \circ \phi)$. Assume that for all $(M, w) \in Mod(\psi)$, it is not the case that $M', w' \leftrightarrows_{C'} M, w$. Then for all $(M, w) \in Mod(\psi)$, $M, w \models \neg \delta_{M'}^{C'}(w')$ by proposition 4.5. So $\psi \rightarrow \neg \delta_{M'}^{C'}(w')$. Then $\psi \circ \phi \rightarrow \neg \delta_{M'}^{C'}(w')$ by application of $(RG1)$. Hence $M', w' \models \neg \delta_{M'}^{C'}(w')$, which is contradictory. ∎

Proposition B.4 (Proposition 4.8)
Revision operator $\circ$ satisfies $(RG2)$ iff for all $\phi \in \mathcal{L}_{\neq Y}^{C}$, for all $(M, w) \in Mod(\psi)$ there is $(M', w') \in Mod(\psi * \phi)$ such that $M, w \leftrightarrows_{C'} M', w'$, with $C' = C - C(\phi)$.

Proof. Similar to Proposition 4.7. ∎

Proposition B.5 (Proposition 4.9)
Revision operator $\circ$ satisfies $(RG1)$ and $(RG2)$ iff for all $\phi \in \mathcal{L}_{\neq Y}^{C}$, $Mod(\psi) \leftrightarrows_{C'} Mod(\psi * \phi)$, with $C' = C - C(\phi)$.

Proof. Follows straightforwardly from Proposition 4.8 and 4.9. ∎

C Proofs of Propositions 5.3, 5.8, 5.9, 5.11, 5.13 and 5.15, and Corollary 5.10

Proposition C.1 (Proposition 5.3)
Let $L = \{(l_0^i, \ldots, l_k^i) \mid i \in S\} \subseteq [0;1]^{k+1}$ and $(l_0^{i_0}, \ldots, l_k^{i_0}) \in L$ (where S is an index set which is possibly infinite).

$$\text{If } (l_0^{i_0}, \ldots, l_k^{i_0}) \geq (l_0^i, \ldots, l_k^i) \text{ for all } i \in S, \text{ then } (l_0^{i_0}, \ldots, l_k^{i_0}) = Sup^k(L).$$

Proof. Let $(A_0, \ldots, A_k) = Sup^k(L)$. We prove by induction on m that $A_m = l_m^{i_0}$.
- $A_k = Sup\{l_k^i \mid i \in S\} = l_k^{i_0}$ by definition of $\leq$.
- Assume for all $k \geq j > m$ that $l_j^{i_0} = A_j$. Then
  $A_m = Sup\{l_m^i \mid l_j^i = A_j$ for all $j > m\}$
  $= Sup\{l_m^i \mid l_j^i = l_j^{i_0}$ for all $j > m\}$ by induction hypothesis
  $= l_m^{i_0}$.

∎

Proposition C.2 (Proposition 5.8)
Let (M, w) be a multi-agent possible world and $\psi \in \mathcal{L}_{\neq Y}$ a satisfiable formula such that $deg(\psi) = d$. Then there is $(M_\psi, w_\psi) \in Mod(\psi)$ such that

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w), R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w_\psi)) \mid j_1, \ldots, j_{d+1} \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1.$$

Proof. We first need to introduce a technical device that will be used in the proof of the proposition.

Definition C.3
Let $d \in \mathbb{N}$. A *tree-like multi-agent possible world of height* d is a finite pointed epistemic model $(M^t, w^t) = (W^t, R^t, V^t, w^t)$ of height d generated by w^t such that:[4]

1. $R_Y(w^t) = \{w^t\}$;
2. for all $j \in G$, R_j is transitive and euclidean;
3. for all $v^t \neq w^t$ there are two unique sequences $v^t_0 = w^t, \ldots, v^t_n = v^t$ and $j_1, \ldots, j_n$ such that $j_i \neq j_{i+1}$, $j_1 \neq Y$ and $w^t = v^t_0 R_{j_1} v^t_1 R_{j_2} \ldots R_{j_n} v^t_n = v^t$;
4. for all v^t and j such that $v^t \in R_j(v^t)$,
   - if $h(v^t) < d$ then for all i, $R_i(v^t) \neq \emptyset$;
   - if $h(v^t) = d$ then for all $i \neq j$, $R_i(v^t) = \emptyset$.

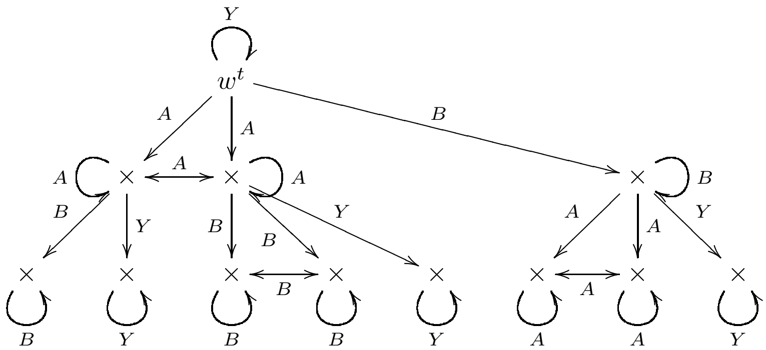

Fig. 12. A tree-like multi-agent possible world of height 2

Now we can prove the proposition.

- One can easily show that there is a tree-like multi-agent possible world of height d, $(M^t, w^t) = (W^t, R^t, V^t, w^t)$, such that:
  – $M^t, w^t \models \psi$
  – for all $j_1, \ldots, j_d$ with $j_i \neq j_{i+1}$, $j_1 \neq Y$ $|R_{j_1} \circ \ldots \circ R_{j_d}(w^t)| \geq |R_{j_1} \circ \ldots \circ R_{j_d}(w)|$.

  For all $j_1, \ldots, j_d$, let $f_{j_1,\ldots,j_d}$ be a surjection from $R_{j_1} \circ \ldots \circ R_{j_d}(w^t)$ to $R_{j_1} \circ \ldots \circ R_{j_d}(w)$.

  For all $j_1, \ldots, j_d$ and $v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w^t)$, we note $M^{v^t} = (W^{v^t}, R^{v^t}, V^{v^t})$ the submodel of M generated by $\bigcup_{i \neq j_d} R_i(f_{j_1,\ldots,j_d}(v^t))$. Then we define $Plug((M^t, w^t), (M, w)) = (W', R', V', w')$ as follows.
  – $W' = W^t \cup \{W^{v^t} \mid v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w^t), j_i \neq j_{i+1}, j_1 \neq Y\}$;
  – $R'_j = R^t_j \cup \{R^{v^t}_j \mid v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w^t), j_i \neq j_{i+1}, j_1 \neq Y\} \cup \{(v^t, u^t) \mid v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w^t), j_i \neq j_{i+1}, j_1 \neq Y, j_d \neq j, u^t \in R_j(f_{j_1,\ldots,j_d}(v^t))(\text{in } M)\}$;
  – $V'(p) = V^t(p) \cup \{V^{f_{j_1,\ldots,j_d}(v^t)}(p) \mid v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w^t), j_i \neq j_{i+1}, j_1 \neq Y\}$;
  – $w' = w^t$.

[4]Let (M, w) be an epistemic model generated by w. The notion of *height* of worlds in M is defined by induction. The only world of height 0 is the root w; the worlds of height $n + 1$ are those immediate successors of worlds of height n that have not yet been assigned a height smaller than $n + 1$. The *height of a (generated) model* (M, w) is the maximum n such that there is a world of height n in (M, w), if such a maximum exists; otherwise the height of (M, w) is infinite.

- Now we prove that $Plug((M^t, w^t), (M, w))$ is a multi-agent possible world. We first prove that $Plug((M^t, w^t), (M, w))$ is serial.
  - For all v such that $h(v) < d$, $R'_j(v) \neq \emptyset$ for all j by condition 4 of the definition of a tree-like multi-agent possible world;
  - for all v such that $h(v) > d$, $R'_j(v) \neq \emptyset$ for all j by definition of a generated submodel;
  - for all v such that $h(v) = d$, $R'_j(v) \neq \emptyset$ for all j by definition of R'_j.

  We prove that condition 2 of the definition of a multi-agent possible world is fulfilled.
  - If $d = 0$ then condition 2 is fulfilled by definition of R'_j;
  - if $d > 0$ then condition 2 is fulfilled by condition 3 of the definition of a tree-like multi-agent possible world.

  The other conditions are obvious.
- Because $deg(\psi) = d$ and the restriction of $Plug((M^t, w^t), (M, w))$ to the worlds of height at most d is bisimilar to (M^t, w^t), we get that $Plug((M^t, w^t), (M, w)), w' \models \psi$.
- Let $v' \in R'_{j_1} \circ \ldots \circ R'_{j_{d+1}}(w')$ with $j_i \neq j_{i+1}$ and $j_1 \neq Y$. Then there is $v^t \in R'_{j_1} \circ \ldots \circ R'_{j_d}(w')$ such that $v' \in R'_{j_{d+1}}(v^t)$. Then $f_{j_1,\ldots,j_d}(v^t) \in R_{j_1} \circ \ldots \circ R_{j_d}(w)$ and there is $v \in R_{j_{d+1}}(f_{j_1,\ldots,j_d}(v^t))$ such that $Plug((M^t, w^t), (M, w)), v' \leftrightarroows M, v$ by definition of $Plug((M^t, w^t), (M, w))$. So $v \in R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w)$ and $M, v \leftrightarroows Plug((M^t, w^t), (M, w)), v'$.

  Likewise, let $v \in R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w)$ with $j_i \neq j_{i+1}$ and $j_1 \neq Y$. Then there is $u \in R_{j_1} \circ \ldots \circ R_{j_d}(w)$ such that $v \in R_{j_{d+1}}(u)$. Then there is $v^t \in R'_{j_1} \circ \ldots \circ R'_{j_d}(w')$ such that $f_{j_1,\ldots,j_d}(v^t) = u$ because $f_{j_1,\ldots,j_d}$ is surjective. Besides, there is $v' \in R'_{j_{d+1}}(v^t)$ such that $Plug((M^t, w^t), (M, w)), v' \leftrightarroows M, v$ by definition of $Plug((M^t, w^t), (M, w))$. So $v' \in R'_{j_1} \circ \ldots \circ R'_{j_{d+1}}(w')$ and $Plug((M^t, w^t), (M, w)), v' \leftrightarroows M, v$.

  So $\sigma(R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w), R'_{j_1} \circ \ldots \circ R'_{j_{d+1}}(w')) = 1$ for all $j_1, \ldots, j_{d+1}$ such that $j_i \neq j_{i+1}$ and $j_1 \neq Y$.

  Therefore $m\{\sigma\left(R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w), R'_{j_1} \circ \ldots \circ R'_{j_{d+1}}(w')\right) \mid j_1, \ldots, j_{d+1} \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1$.
- Finally, we define (M_ψ, w_ψ) as the bisimulation contraction of $Plug((M^t, w^t), (M, w))$. Then all the results for $Plug((M^t, w^t), (M, w))$ still hold for (M_ψ, w_ψ) and besides $(M_\psi, w_\psi) \in Mod(\psi)$. ∎

PROPOSITION C.4 (Proposition 5.9)
Let (M, w) be a multi-agent possible world. For all $k \in \mathbb{N}^*$, there are finitely many multi-agent possible worlds (M', w') such that

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_k}(w'), R_{j_1} \circ \ldots \circ R_{j_k}(w)) \mid j_1, \ldots, j_k \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1.$$

PROOF. We first prove a lemma.

LEMMA C.5
Let $\mathcal{M} = \{(M^1, w^1), \ldots, (M^n, w^n)\}$ be an internal model for agent j.[5]

For all $k \in \mathbb{N}$, there are finitely many multi-agent possible worlds (M', w') for agent j such that

(∗) for all $j_1, \ldots, j_k$ with $j_1 \neq j$ and $j_i \neq j_{i+1}$, for all $v \in R_{j_1} \circ \ldots \circ R_{j_k}(w')$, there is $(M^i, w^i) \in \mathcal{M}$ and $v^i \in R_{j_1} \circ \ldots \circ R_{j_k}(w^i)$ such that $M, v \leftrightarroows M^i, v^i$.

PROOF. First, note that every multi-agent possible world (M', w') for agent j can be seen as the 'connection' of an interpretation (the root w) with a finite number of multi-agent possible worlds for each agent $l \neq j$.

Now we prove the lemma by induction on k.

k=1 Because Φ is finite, there are finitely many interpretations. So there are finitely many (valuations for the) roots of multi-agent possible worlds.

By (∗), there are also finitely many worlds accessible from each root modulo bisimulation. So, by the remark at the beginning of this proof, there are finitely many multi-agent possible worlds satisfying (∗).

---

[5]An internal model or a multi-agent possible world *for agent* j is an internal model or a multi-agent possible world where the designated agent is j instead of Y.

k+1 For all $l \neq j$, for all $(M^i, w^i) \in \mathcal{M}$, let M^i_l be the submodel of M^i generated by $R_l(w^i)$. M^i_l can be seen as the epistemic model associated to an internal model (for agent l). Let $\mathcal{M}^i_l = \{(M^1_l, w^1_l), \ldots, (M^{n_i}_l, w^{n_i}_l)\}$ be this internal model. Let $\mathcal{M}_l = \bigcup_{i \in \{1,\ldots,n\}} \mathcal{M}^i_l = \bigcup_{i \in \{1,\ldots,n\}} \{(M^1_l, w^1_l), \ldots, (M^{n_i}_l, w^{n_i}_l)\}$.

Now, using the remark at the beginning of this proof,

there are finitely many multi-agent possible worlds for agent j satisfying (*)

iff for all $l \neq j$ there are finitely many multi-agent possible worlds (M', w') for agent l such that for all $j_1, \ldots, j_k$ with $j_1 \neq l$ and $j_i \neq j_{i+1}$,
for all $v' \in R_{j_1} \circ \ldots \circ R_{j_k}(w')$, there is $(M^i, w^i) \in \mathcal{M}$ and $v^i \in R_l \circ R_{j_1} \circ \ldots \circ R_{j_k}(w^i)$ such that $M', v' \leftrightarrows M^i, v^i$.

iff for all $l \neq j$ there are finitely many multi-agent possible worlds (M', w') for agent l such that for all $j_1, \ldots, j_k$ with $j_1 \neq l$ and $j_i \neq j_{i+1}$,
for all $v' \in R_{j_1} \circ \ldots \circ R_{j_k}(w')$, there is $(M^i_l, w^i_l) \in \mathcal{M}_l$ and $v^i_l \in R_{j_1} \circ \ldots \circ R_{j_k}(w^i_l)$ such that $M', v' \leftrightarrows M^i_l, v^i_l$,
which is true by induction hypothesis.

∎

The proof follows easily from the lemma. Indeed, we just take $\mathcal{M} = \{(M, w)\}$ and we can then apply the lemma because for all $k \in \mathbb{N}^*$ and all multi-agent possible worlds (M', w'), if

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_k}(w'), R_{j_1} \circ \ldots \circ R_{j_k}(w)) \mid j_1, \ldots, j_k \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1$$

then $(*)$ is fulfilled. ∎

Corollary C.6 (Corollary 5.10)
Let (M, w) be a multi-agent possible world, $\psi \in \mathcal{L}_{\neq Y}$ and $k = deg(\psi) + 1$. Then there is $(M', w') \in Mod(\psi)$ such that $s^k((M', w'), (M, w)) = s^k(Mod(\psi), \{(M, w)\})$.

Proof. It follows from Propositions 5.8 and 5.9. ∎

Proposition C.7 (Proposition 5.11)
The assignment defined in Definition 5.7 is a faithful assignment. Therefore the operator $\circ$ defined in Definition 5.7 satisfies the postulates $(R1) - (R6)$. Besides, $\circ$ satisfies also postulates $(RG1)$.

Proof. • Clearly $\leq_\psi$ is a total pre-order because $\leq^k$ is a total pre-order. We are going to show that it is faithful.

- If $(M, w), (M', w') \in Mod(\psi)$ then $s^k(Mod(\psi), (M, w)) = s^k(Mod(\psi), (M', w')) = (1, \ldots, 1)$ by definition of s^k. So we cannot have $(M, w) <_\psi (M', w')$.
- If $(M, w) \in Mod(\psi)$ and $(M', w') \notin Mod(\psi)$ then $s^k(Mod(\psi), (M, w)) = (1, \ldots, 1)$ and $s^k(Mod(\psi), (M', w')) = (l_1, \ldots, l_k)$ with $l_1 < 1$.
  So $s^k(Mod(\psi), (M, w)) >^k s^k(Mod(\psi), (M', w'))$, i.e. $(M, w) <_\psi (M', w')$.
- Finally, if $\psi \leftrightarrow \psi'$ then clearly $\leq_\psi = \leq_{\psi'}$.

• We are going to show that $\circ$ satisfies postulate $(RG1)$. Let $\phi \in \mathcal{L}_{\neq Y}$ and $(M', w') \in Mod(\psi \circ \phi)$. Assume that for all $(M, w) \in Mod(\psi)$, it is not the case that $M, w \leftrightarrows_{C'} M', w'$ with $C' = C_0 - C(\phi)$.

Let $(M, w) \in Mod(\psi)$ such that $s^k((M, w), (M', w')) = s^k(Mod(\psi), (M', w'))$. Such a (M, w) exists by Corollary 5.10.
Assume that `pf` $\notin C'$, the case `pf` $\in C'$ is dealt with similarly. Then by definition of $\leftrightarrows_{C'}$,
there is $j_0 \in C'$, $v \in R_{j_0}(w)$ such that for all $v' \in R_{j_0}(w')$ it is not the case that $M, w \leftrightarrows M', v'$ (1)
or there is $j_0 \in C'$, $v' \in R_{j_0}(w')$ such that for all $v \in R_{j_0}(w)$ it is not the case that $M, v \leftrightarrows M', v'$ (2).
Assume w.l.o.g. that (1) is the case. Then $\sigma(R_{j_0}(w), R_{j_0}(w')) < 1$.
Using generated submodels, we can easily build a multi-agent possible world (M'', w'') such that $M'', w'' \leftrightarrows_{C_0 - \{j_0\}} M', w'$ and $M'', w'' \leftrightarrows_{j_0} M, w$.

- Then for all $j \neq j_0$, $\sigma(R_j(w''), R_j(w)) = \sigma(R_j(w'), R_j(w))$ and $\sigma(R_{j_0}(w''), R_{j_0}(w)) = 1 > \sigma(R_{j_0}(w', R_{j_0}(w))$.
- So for all $n \in \mathbb{N}^*$, all $j_1, \dots, j_n$ such that $j_1 \neq j_0, j_i \neq j_{i+1}$, $\sigma(R_{j_0} \circ R_{j_1} \circ \dots \circ R_{j_n}(w''), R_{j_0} \circ R_{j_1} \circ \dots \circ R_{j_n}(w)) = 1 \geq \sigma(R_{j_0} \circ R_{j_1} \circ \dots \circ R_{j_n}(w'), R_{j_0} \circ R_{j_1} \circ \dots \circ R_{j_n}(w))$.
  Besides, for all $n \in \mathbb{N}^*$, all $j_1, \dots, j_n$ such that $j_1 \neq j_0$, $j_1 \neq Y$, $j_i \neq j_{i+1}$, $\sigma(R_{j_1} \circ \dots \circ R_{j_n}(w''), R_{j_1} \circ \dots \circ R_{j_n}(w)) = \sigma(R_{j_1} \circ \dots \circ R_{j_n}(w'), R_{j_1} \circ \dots \circ R_{j_n}(w))$ because $M'', w'' \leftrightarrows_{C_0 - \{j_0\}} M', w'$.

So for all $n \geq 2$, $m\{\sigma(R_{j_1} \circ \dots \circ R_{j_n}(w''), R_{j_1} \circ \dots \circ R_{j_n}(w)) \mid j_i \neq j_{i+1}, j_1 \neq Y\} \geq m\{\sigma(R_{j_1} \circ \dots \circ R_{j_n}(w'), R_{j_1} \circ \dots \circ R_{j_n}(w)) \mid j_i \neq j_{i+1}, j_1 \neq Y\}$ and $m\{\sigma(R_j(w''), R_j(w)) \mid j \in G, j \neq Y\} > m\{\sigma(R_j(w'), R_j(w)) \mid j \in G, j \neq Y\}$.

So $s^k((M'', w''), (M, w)) >^k s^k((M', w'), (M, w))$.

Finally, because $\phi \in \mathcal{L}_{C_0 - \{j_0\}}$, $M'', w'' \leftrightarrows_{C_0 - \{j_0\}} M', w'$ and $M', w' \models \phi$, we have $M'', w'' \models \phi$. So $(M'', w'') \in Mod(\phi)$. Then $(M', w') \notin Mod(\psi \circ \phi)$ which is impossible by assumption. ∎

Proposition C.8 (Proposition 5.13)
Let $\mathcal{M}$ be an internal model and $\phi \in \mathcal{L}_{\neq Y}$ a satisfiable formula. Then $\mathcal{M} * \phi$ is an internal model.

Proof. Let $k = deg(\phi) + 1$. By Proposition 5.8, we know that there is $(M', w') \in Mod(\phi)$ and $(M, w) \in \mathcal{M}$ such that

$$m\{\sigma(R_{j_1} \circ \dots \circ R_{j_k}(w'), R_{j_1} \circ \dots \circ R_{j_k}(w)) \mid j_1, \dots, j_k \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1. (**)$$

So $\mathcal{M} * \phi = \{(M', w') \in Mod(\phi) \mid s^k((M', w'), \mathcal{M}) = s^k(Mod(\phi), \mathcal{M})\} = \{(M', w') \in Mod(\phi) \mid$ there is $(M, w) \in \mathcal{M}$ such that (M', w') satisfies (**) and $s^k((M', w'), (M, w)) = s^k(Mod(\phi), \mathcal{M})\}$. By proposition 5.9, this last set is finite. So $\mathcal{M} * \phi$ is finite, i.e. $\mathcal{M} * \phi$ is an internal model. ∎

Proposition C.9 (Proposition 5.15)
$\{(M, w), (M', w')\} * B_A p = \{(M^r, w^r), (M^{r'}, w^{r'})\}$

Proof. We first prove a series of lemmas.

Lemma C.10
Let (M'', w'') such that $M'', w'' \models B_A p$ and $s^2(\{(M, w), (M', w')\}, (M'', w'')) = s^2(\{(M, w), (M', w')\}, Mod(B_A p))$. Then $|M''| \geq 4$.

Proof. W.l.o.g. we assume that $M'', w'' \models p$. Let $v'' \in R_A \circ R_Y(w'')$. We know by Proposition 5.8 that $s^2(\{(M, w), (M', w')\}, (M'', w'')) = (\alpha, \beta, 1)$. So there is $v \in R_A \circ R_Y(w)$ such that $M'', v'' \leftrightarrows M, v$.

Then there are $v''_1, v''_2 \in R_A \circ R_Y(w'')$ such that $M'', v''_1 \models p \wedge \neg B_A p \wedge \neg B_Y p$ and $M'', v''_2 \models \neg p \wedge \neg B_A p \wedge \neg B_Y p$. There is also $v''_3 \in R_A(w'')$ such that $M'', v''_3 \models B_A p \wedge \neg B_Y p$. Finally, $M'', w'' \models B_Y p \wedge B_A p$. So we have 4 worlds w'', v''_1, v''_2 and v''_3 satisfying different formulas. Therefore, there are at least 4 worlds in M''. ∎

Lemma C.11
Let (M'', w'') such that $M'', w'' \models B_A p$. Then

$$s^2(\{(M, w), (M', w')\}, (M'', w'')) = \left(\frac{1}{3|M''| + 1}, \frac{3}{4(3|M''| + 1)}, \alpha\right)$$

for some $\alpha \in [0; 1]$.

Therefore $s^2(\{(M, w), (M', w')\}, Mod(B_A p)) \leq (\frac{1}{13}, \frac{3}{52}, 1)$.

Proof. Assume w.l.o.g. that $M'', w'' \models p$. Then $max\{i \mid M'', w'' \leftrightarrows_i M, w\} = 1$.

Let $v'' \in R_A(w'')$. Then $max\{i \mid M'', v'' \leftrightarrows_i M, v$ and $v \in R_A(w)\} = 1$ because $M'', v'' \models B_A p$ and $M, v_i \not\models B_A p$ for $i = 1, 2$.

$max\{i \mid M'', v'' \leftrightarrows_i M, v_1$ and $v'' \in R_A(w'')\} = 1$ because for all $v'' \in R_A(w'')$, $M'', v'' \models B_A p$ and $M, v_i \not\models B_A p$ for $i = 1, 2$.

$max\{i \mid M'', v'' \leftrightarrows_i M, v_2 \text{ and } v'' \in R_A(w'')\} = 0$ because for all $v'' \in R_A(w'')$, $M'', v'' \models p$ and $M, v_2 \models \neg p$.

So $\sigma(w, w') = \frac{1}{|M||M''|+1}$, and

$\sigma(R_A(w), R_A(w'')) = \frac{1}{2}\left(\frac{1}{2}\left(\frac{1}{|M||M''|+1} + \frac{0}{|M||M''|+1}\right) + \frac{1}{|R_A(w'')|}\sum_{v'' \in R_A(w'')} \frac{1}{|M||M'|+1}\right) =$

$\frac{1}{2}\left(\frac{1}{2}\frac{1}{|M||M''|+1} + \frac{1}{|M||M''|+1}\right) = \frac{3}{4(|M||M''|+1)}$.

Therefore $s^2(\{(M, w), (M', w')\}, (M'', w'')) = \left(\frac{1}{3|M''|+1}, \frac{3}{4(3|M''|+1)}, \alpha\right)$ for some $\alpha \in [0; 1]$.

We get that $s^2(\{(M, w), (M', w')\}, Mod(B_A p)) \leq \left(\frac{1}{13}, \frac{3}{52}, 1\right)$ thanks to Lemma C.10. ∎

Lemma C.12

$s^2(\{(M, w), (M', w')\}, (M^r, w^r)) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$.
$s^2(\{(M, w), (M', w')\}, (M^{r'}, w^{r'})) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$.

Lemma C.13
Let (M'', w'') such that $M'', w'' \models B_A p$. Then,
if $s^2((M, w), (M'', w'')) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$ then $M'', w'' \leftrightarrows M^r, w^r$;
if $s^2((M', w'), (M'', w'')) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$ then $M'', w'' \leftrightarrows M^{r'}, w^{r'}$.

Proof. Assume $s^2\left((M, w), (M'', w'')\right) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$. Then $|M''| = 4$ by Lemma C.11. Then one can easily show that $|R_A(w'')| = 1$ and $|R_A \circ R_Y(w'')| = 2$. We set $R_A(w'') = \{v_1\}$ and $R_A \circ R_Y(w'') = \{v_3'', v_4''\}$ with $M'', v_3'' \models p$ and $M'', v_4'' \models \neg p$.

Let $Z = \{(w^r, w''), (v_2^r, v_2''), (v_3^r, v_3''), (v_4^r, v_4'')\}$. One can easily show that Z is a bisimulation between (M^r, w^r) and (M'', w'').

The proof is similar if $s^2\left((M', w'), (M'', w'')\right) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$. ∎

The proof of Proposition 5.15 then follows easily from Lemma C.11, Lemma C.12 and Lemma C.13. ∎

Received December 2007