Communication among agents: a way to revise beliefs in KD45 Kripke structures*

Jean-Marc Tallon* — **Jean-Christophe Vergnaud***
Shmuel Zamir* ****

* *EUREQua, CNRS,*
Université Paris I Panthéon-Sorbonne
{jmtallon,vergnaud}@univ-paris1.fr

** *CREST–LEI*
and Center for the Study of Rationality,
Hebrew University
zamir@math.huji.ac.il

ABSTRACT. *We address the issue of belief revision in a multi-agent setting. We represent agents' beliefs in a semantic manner, through a Kripke structure, and model a communication process by which agents communicate their beliefs to one another. We define a revision rule that can be applied even when agents have contradictory beliefs. We study its properties and show that agents need not agree after communicating their beliefs. We finally address the dynamics of revision and show that the order of communication may affect the resulting belief structure.*

KEYWORDS: *belief revision, KD45, Kripke structure.*

1. Introduction

Situations in which agents have mistaken beliefs abound. In this paper, we propose a revision rule that specifies how agents' beliefs evolve after communication among themselves has taken place. Specifically, we work with KD45 Kripke structures (e.g., [CHE 80]) and allow agents to communicate (non strategically) their beliefs. The issue is to come up with a rule specifying how initial beliefs that are contradicted by the announcement of some other agent are changed to cope with this contradiction.

* We would like to thank J. Lang, B. Walliser, and an anonymous referee for useful comments. We also thank the audience at RUD2002 and participants at the workshop on Belief Revision, LIP6, U. Paris VI. Financial support from the French Ministry of Research (ACI Cognitique) is gratefully acknowledged.

In the absence of any mistake (i.e., in S5), the process is simple: each agent simply drops from his beliefs the ones that are not compatible with the announcements. This yields a new S5 Kripke structure. However, in the presence of mistaken initial beliefs, the rule has to propose a way to correct these beliefs. We proceed in two steps. First, we specify a selection procedure that specifies which initial beliefs the agent retains upon hearing the other's announcement. We do so by defining agent selection functions on the possible worlds. These selection functions allow for the possibility that agents eliminate certain of their initial beliefs. We do not try to ground these selection functions on some rational basis and leave them essentially unconstrained. The only restriction we consider is a minimal consistency requirement which says the following: any state, initially believed possible, that is not contradicted by the announcements is still considered possible after the revision process. We also modify the accessibility relation so that the beliefs of the agents who have announced are now commonly known. We show that this rule is always well defined, in the sense that it leads to a KD45 Kripke structure. We provide conditions under which no revision occurs. We also give conditions on the initial structure that guarantee the emergence of consensus. We next extend this rule to a dynamic setting, in which agents announce and revise their beliefs sequentially. We show that, interestingly, the rule proposed is commutative whenever agents' beliefs are correct (that is, in S5), but that in general, in KD45, the order according to which the agents announce their beliefs might matter. For instance, the final epistemic situation reached is not the same whether all agents announced simultaneously or one at a time.[1]

We consider only a semantic framework. In certain fields such as economics and game theory, the semantic approach is favored while logicians rather prefers to consider the syntax. We will discuss informally along the text the issue of belief revision from a syntaxical point of view. In particular, we will discuss the difficulties about stating some axioms of belief revision in a multi agents situation with respect to the AGM axioms for a single agent's setting. Indeed, while it would be nice to adapt the AGM [ALC 85] axiomatic method to multi-agent belief revision, this is far from being a simple task.[2] Indeed, belief revision in a multi-agent framework poses not only the problem of integrating new information but also the issue of how agents perceive how other agents will integrate new information. This could lead for instance to violations of AGM's axiom of success.

2. Minimal Kripke structures: definition and preliminaries

Let $I = \{1, ..., i, ..., n\}$ be a finite set of agents and S a set of states of nature (for instance the game that is being played among the agents). A Kripke structure is a representation of agents' beliefs about the state of nature s and about the beliefs of the other agents.

1. The possibility of non commutativity of belief revision has already been noticed in the literature. See [GER 97].
2. For an attempt in that direction, see [BOA 03].

DEFINITION 1. — *A Minimal Kripke Structure (MKS) is a collection* $(\Omega, \omega_0, s, (t_i)_{i\in I})$*, where* Ω *is a set, and the following conditions are satisfied:*

(i) s *is a mapping from* Ω *to* S*;*

(ii) $\forall i \in I, t_i$ *is a mapping from* Ω *to* 2^{Ω}*;*

(iii) $\forall i \in I, \forall \omega \in \Omega, \omega' \in t_i(\omega) \Rightarrow t_i(\omega') = t_i(\omega)$*;*

(iv) $\omega_0 \in \Omega$*;*

(v) there does not exist $\Omega' \subsetneq \Omega$ *such that* $(\Omega', \omega_0, s|_{\Omega'}, (t_i|_{\Omega'})_{i\in I})$ *satisfies conditions (i) to (iv).*[3]

We will refer to an element $(\omega; s(\omega); t_1(\omega), ..., t_n(\omega))$ as a *state of the world.* ω is the name of the state, $s(\omega)$ is the state of nature in the world ω, $t_i(\omega)$ is the set of states of the world that i considers possible in state ω. Finally, ω_0 is the true state of the world. Abusing notation slightly we will denote a state of the world $\omega = (s(\omega), t_1(\omega), ..., t_n(\omega))$.

Since we do not require that agents consider ω_0 possible, the structures we look at may contain mistaken beliefs. Hence, we place ourselves in the system KD45 rather than S5. Embedded in the definition are several assumptions about the nature of the situations we model. First, we assume a form of consistency of the beliefs: (iii) of the definition implies that beliefs are partitional (i.e., $\{t_i(\omega)\}_{\omega\in\Omega}$ is a partition of $\Omega_i =: \cup_{\omega\in\Omega} t_i(\omega)$). Note however that Ω_i is not necessarily equal to Ω. Second, the true state ω_0 is given, since by construction an MKS is a representation of given beliefs (the ones encapsulated in ω_0, the other states being part of the description of these beliefs). Third, we assume that the Kripke structure is minimal in the sense that it does not contain a smaller Kripke structure (condition (v)). This last condition is equivalent to assuming that the system does not contain states that are not deemed possible via a finite sequence of steps of the form "I think that you think that she thinks..." (condition (v') in Proposition 2 below, which will be used repeatedly in the proofs of this paper.) This does not imply that Ω is finite.

PROPOSITION 2. — [4] *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be a collection which satisfies conditions (i) to (iv) of Definition 1. Then condition (v) is equivalent to*

(v') $\forall \omega \in \Omega \setminus \{\omega_0\}$, *there exists a finite sequence,* $\{i_k\}_{k=1}^{k=r}$ *with* $i_k \in I$ *for all* k *such that* $\omega \in t_{i_1}(t_{i_2}(...(t_{i_r}(\omega_0))))$ *where for any* $A \subseteq \Omega$, $t_i(A) = \cup_{\omega\in A} t_i(\omega)$.

Condition (v') defines what is often called the generated sub-model. The classical example of the muddy children can be expressed in this formalism.

EXAMPLE 3. — Three children come home after playing in the field. They might have a clean face (C) or a dirty face (D). Each child sees the other two's faces but does not see whether there is mud on her own face. Assume that the three faces are actually dirty. Denoting the states of nature by $F_1F_2F_3$ where $F_i \in \{C, D\}$ is the state of i's face, we represent this situation by an MKS given by $\Omega = \{\omega_0, ..., \omega_7\}$ where,

---

3. $t_i|_{\Omega'}$ is the restriction of t_i to Ω', i.e., $t_i|_{\Omega'} : \Omega' \to 2^{\Omega}$ and $t_i|_{\Omega'}(\omega) = t_i(\omega)$ for all $\omega \in \Omega'$.
4. All proofs are gathered in Appendix B.

– $\omega_0 = (DDD, \{\omega_0, \omega_4\}, \{\omega_0, \omega_2\}, \{\omega_0, \omega_1\})$
– $\omega_1 = (DDC, \{\omega_1, \omega_5\}, \{\omega_1, \omega_3\}, \{\omega_0, \omega_1\})$
– $\omega_2 = (DCD, \{\omega_2, \omega_6\}, \{\omega_0, \omega_2\}, \{\omega_2, \omega_3\})$
– $\omega_3 = (DCC, \{\omega_3, \omega_7\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\})$
– $\omega_4 = (CDD, \{\omega_0, \omega_4\}, \{\omega_4, \omega_6\}, \{\omega_4, \omega_5\})$
– $\omega_5 = (CDC, \{\omega_1, \omega_5\}, \{\omega_5, \omega_7\}, \{\omega_4, \omega_5\})$
– $\omega_6 = (CCD, \{\omega_2, \omega_6\}, \{\omega_4, \omega_6\}, \{\omega_6, \omega_7\})$
– $\omega_7 = (CCC, \{\omega_3, \omega_7\}, \{\omega_5, \omega_7\}, \{\omega_6, \omega_7\})$ □

The next example illustrates an instance of mistaken beliefs.

EXAMPLE 4. — Let $S = \{\alpha, \beta\}$, $I = \{1, 2\}$ and $\Omega = \{\omega_0, \omega_1, \omega_2, \omega_3\}$ such that:

– $\omega_0 = (\alpha, \{\omega_1, \omega_2\}, \{\omega_3\})$
– $\omega_1 = (\alpha, \{\omega_1, \omega_2\}, \{\omega_1, \omega_2\})$
– $\omega_2 = (\beta, \{\omega_1, \omega_2\}, \{\omega_1, \omega_2\})$
– $\omega_3 = (\beta, \{\omega_3\}, \{\omega_3\})$ □

To describe the situation which is represented in this structure, let us introduce some elements of syntax. First, for notational simplicity we will denote also by α, β... the primitive propositions i.e: considered as a proposition, α means "the nature is in state α". We note $\wedge, \vee, \neg$, and $\rightarrow$ for respectively, the and, or, negation and material implication operators. We consider individual belief operators b_i and a common belief operator cb. [5] Therefore, the previous example catches a situation where the proposition $\alpha \wedge b_1 cb (\alpha \vee \beta) \wedge b_2 cb\beta$ holds true, that is, in words, a situation where the state of nature is α, where agent 1 believes that it is common belief that α or β and agent 2 believes that it is common belief that β.

A given epistemic situation could be captured by MKS that are formally different. This fact is not bothersome in S5, i.e., if agents do not make any mistake. However, as we want to study revision in beliefs when agents potentially have initial mistaken beliefs, we have to make sure that "irrelevant" mistakes can be dropped at the outset so as to focus on beliefs that are mistaken in a meaningful way. A simple intuition of why some mistakes are not meaningful is the following: imagine that $\omega_0 \notin t_i(\omega_0)$. This can reflect two very different situations: either the agent is correct in the sense that in ω_0 he believes possible a state ω' which represents the same beliefs as ω_0; or the agent is making a mistake in the sense that he is not considering as possible the true state of the world ω_0 or any state of the world that represents the same epistemic state.

5. The common belief operator cb has the intuitive meaning that everybody believes that everybody believes...an infinite number of time. Since we do not allow for infinite conjunction, the definition of the common belief operator cannot be defined from the individual belief operators and its properties have to be defined per se. For instance, we have $cb\varphi \rightarrow b_i\varphi$ and $cb\varphi \rightarrow b_i cb\varphi$ for all i and proposition φ. See [BON 96].

EXAMPLE 5. — Let $S = \{\alpha\}$, $I = \{1, 2\}$ and consider two MKS.
$\Omega = \{\omega_0, \omega_1\}$ such that:

- $\omega_0 = (\alpha, \{\omega_1\}, \{\omega_1\})$
- $\omega_1 = (\alpha, \{\omega_1\}, \{\omega_1\})$

and $\Omega' = \{\omega'_0\}$ such that:

- $\omega'_0 = (\alpha, \{\omega'_0\}, \{\omega'_0\})$

These two MKS represent the same situation: α is the state of nature and it is common belief that α is the state of nature. □

A way of getting around this difficulty is to define notions of *representation* and *equivalence* of MKS as well as a notion of *irreducibility* for MKS.

DEFINITION 6. — *An MKS,* $(\Omega', \omega'_0, s', (t'_i)_{i\in I})$*, is a* representation *of the MKS* $(\Omega, \omega_0, s, (t_i)_{i\in I})$*, if there exists a mapping* σ *from* Ω *to* Ω' *such that*

(i) $\sigma(\Omega) = \Omega'$

(ii) $\sigma(\omega_0) = \omega'_0$

(iii) $s' \circ \sigma = s$

(iv) $\forall i \in I, t'_i \circ \sigma = \sigma \circ t_i$.

DEFINITION 7. — *Two MKS,* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *and* $(\Omega', \omega'_0, s', (t'_i)_{i\in I})$*, are* equivalent *if they have a common representation,* $(\Omega'', \omega''_0, s'', (t''_i)_{i\in I})$.

This notion of equivalence corresponds to bisimulation. We now define a notion of redundancy within an MKS.

DEFINITION 8. — *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MKS. Two states* $\omega_1, \omega_2 \in \Omega$ *are said to be* identical *if there exists an MKS,* $(\Omega', \omega'_0, s', (t'_i)_{i\in I})$ *and a mapping* $\sigma : \Omega \to \Omega'$ *as in Definition 6 such that* $\sigma(\omega_1) = \sigma(\omega_2)$.

Two states of the world are thus identical if there exists a representation of the MKS in which these two states are represented by the same state of the world. Our next step is to define irreducible MKS, in which such a problem does not arise.

DEFINITION 9. —

– *An MKS,* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *is* irreducible *if no two distinct states of the world* $\omega, \omega' \in \Omega$*, are identical.*

– *An MKS,* $(\Omega', \omega'_0, s', (t'_i)_{i\in I})$ *is an* irreducible representation *of* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *if it is a representation of* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *and it is irreducible.*

In the paper we deal exclusively with irreducible MKS. This is without loss of generality as the next proposition makes it clear, since non irreducible MKS always have an irreducible representation.

PROPOSITION 10. — *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MKS. Then it has an irreducible representation* $(\Omega', \omega'_0, s', (t'_i)_{i\in I})$ *and all its irreducible representations are equivalent.*

Finally, we define a notion of correctness for MKS.

DEFINITION 11. — *Let $(\Omega, \omega_0, s, (t_i)_{i\in I})$ be an irreducible MKS. An agent $i \in I$ has* correct beliefs *if $\omega_0 \in t_i(\omega_0)$. The MKS is* correct *if all agents have correct beliefs. The MKS is* totally correct *if $\omega \in t_i(\omega)$ for all $\omega \in \Omega$ and all $i \in I$.*[6]

Total correctness amounts to assume S5. Obviously, an MKS can be correct but not totally correct, as illustrated in the following example.

EXAMPLE 12. — Let $S = \{\alpha, \beta\}$ and $I = \{1, 2\}$. Consider $\Omega = \{\omega_0, \omega_1\}$ where

– $\omega_0 = (\alpha, \{\omega_0\}, \{\omega_0, \omega_1\})$
– $\omega_1 = (\beta, \{\omega_0\}, \{\omega_0, \omega_1\})$ □

3. Common belief in minimal Kripke structures

When agents hold mistaken beliefs, they do not necessarily all have the same view of what the model actually is. We introduce here the notion of belief horizon of an agent which is the model the agent has in mind.

DEFINITION 13. — *Let $(\Omega, \omega_0, s, (t_i)_{i\in I})$ be an MKS. The* belief horizon *of agent $i \in I$, denoted by $BH_i(\omega_0, t)$, is the minimal subset Y of Ω satisfying:*

(i) $t_i(\omega_0) \subseteq Y$,
(ii) $\forall \omega \in Y, \forall j \in I, t_j(\omega) \subseteq Y$.

Thus, $BH_i(\omega_0, t)$ is the smallest set such that i believes it and believes that all other agents believe it, believes that others believe that others believe it and so forth. In Example 4, one has $BH_1(\omega_0, t) = \{\omega_1, \omega_2\}$ and $BH_2(\omega_0, t) = \{\omega_3\}$.

PROPOSITION 14. — *Let $(\Omega, \omega_0, s, (t_i)_{i\in I})$ be an MKS. Then,*

$$\Omega = \{\omega_0\} \cup \left(\bigcup_{i\in I} BH_i(\omega_0, t) \right)$$

Define now the notion of common belief.

DEFINITION 15. — *Let $(\Omega, \omega_0, s, (t_i)_{i\in I})$ be an MKS. An event $E \subseteq \Omega$ is* common belief *(CB) if for any $r \in \mathbb{N}$ and sequence $\{i_k\}_{k=1}^{k=r}$, $i_k \in I$, $t_{i_1}(t_{i_2}(\ldots(t_{i_r}(\omega_0))) \subseteq E$.*

Note that as an MKS describes a mutual belief structure at a specific, "true", state of the world, common belief is also defined at that state ω_0. The following proposition characterizes those events that are common beliefs.

6. If the MKS considered were not irreducible, the definition should be slightly more general: an MKS is correct if $\forall i \in I$, there exists $\omega \in t_i(\omega_0)$, such that ω and ω_0 are identical. When the MKS is irreducible, this definition and Definition 11 coincide.

PROPOSITION 16. — *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MKS. An event* $E \subseteq \Omega$ *is common belief if and only if* $BH_i(\omega_0, t) \subseteq E$ *for all* $i \in I$.

This notion of common belief is meaningful for the analyst since, according to i's beliefs, *any event* containing $BH_i(\omega_0, t)$ is CB. As we shall see later, only at the absence of mistakes, CB events have stronger meaning.

COROLLARY 17. — *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MKS. An event* $E \subseteq \Omega$ *is common belief if and only if*

$$\cup_{i\in I} BH_i(\omega_0, t) \subseteq E \subseteq \Omega = \{\omega_0\} \cup \left(\bigcup_{i\in I} BH_i(\omega_0, t)\right)$$

This corollary establishes that in an MKS, at most two events can be common belief. Ω is always commonly believed (by construction), while $\Omega \setminus \{\omega_0\}$ is common belief only if the true state ω_0 does not belong to the belief horizon of any agent. In other words, Ω is the only common belief event at ω_0 if and only if ω_0 is in the belief of at least one agent, that is, if and only if there exists i such that $\omega_0 \in t_i(\omega_0)$.

In syntaxical terms, that means that $cb\varphi$ is true (in the real state ω_0) if and only if φ is true in all the worlds of $BH_i(\omega_0, t)$ for all i.

4. Communication and revision in minimal Kripke structures

We are interested in studying the evolution of beliefs when agents can communicate their beliefs to each other and update accordingly. In this section we provide a rule according to which agents revise their beliefs in a communication process. At this stage of our work, we do not allow agents to announce false (or partly false) or even imprecise beliefs. Thus, the analysis will concentrate on the case in which agents announce *truthfully and precisely* their beliefs.

DEFINITION 18. — *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MKS. A* communication *is simply a subset* I^c *of* I, *of agents that announce their beliefs (i.e.,* $(t_i(\omega_0))_{i\in I^c}$*).*

A communication can be identified by $I^c \subseteq I$, the group of agents who announce their true beliefs. We'll refer to it as *full communication* when $I^c = I$. The restriction that agents announce precisely their true beliefs can be understood as an assumption that the information revealed can be somehow certified. We will assume in the sequel that it is "common belief" that agents announce precisely their true beliefs. For instance, in Example 4, full communication by agent 1 and 2 means concretely that agent 1 announces publicly that he believes that the state of nature is α or β, that he believes that this is common belief, while agent 2 announces that he believes it is common belief that the state of nature is β.