Contradicting Beliefs and Communication : the Static Case[1]

Jean-Marc Tallon† jmtallon@univ-paris1.fr
Jean-Christophe Vergnaud† vergnaud@univ-paris1.fr
Shmuel Zamir‡ zamir@ensae.fr

†CNRS–EUREQua

‡CNRS–EUREQua et CREST–LEI

Abstract:
We address the issue of the representation as well as the evolution of (possibly) mistaken beliefs. We develop a formal setup (a mutual belief space) in which agents might have a mistaken view of what the model is. We then 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, in particular, show that, when mistaken, agents do not necessarily eventually agree after communicating their beliefs. We finally address the dynamics of revision and show that when beliefs are mistaken, the order of communication may affect the resulting belief structure.
Keywords : Mutual beliefs system, Communication, Revision

Résumé :

Nous étudions la représentation et l'évolution de croyances, potentiellement erronées. Nous développons un cadre formel dans lequel les agents peuvent se tromper sur ce qu'est le vrai modèle. Nous introduisons alors un processus de communication, au travers duquel les agents apprennent les croyances des autres. Nous développons une règle de révision des croyances qui s'applique également dans le cas où les croyances des agents sont contradictoires entre elles. Nous étudions les propriétés de cete règle et montrons que, lorsque les croyances initiales sont erronées, un désaccord entre agents peut persister même après communication. Nous traitons enfin de la dynamique de cette règle et montrons que l'ordre des annonces peut dans certains cas affecter la structure finale des croyances.
Mots clé : Système de croyances mutuelles, Communication, Révision

1 Introduction

In this paper, we address the issue of the representation as well as the evolution of (possibly) mistaken beliefs. The formal setup we develop allows one to model situations in which agents do not have the same view of what is the actual model of the economy. Hence, after communication of each other's beliefs, they might have to deal with surprises or unforeseen contingencies : agent i may be proven wrong in his beliefs about what j thinks the model is after he (i) hears j's announcement. As a consequence, a lot of the intuition one has formed in the standard case (i.e., with no mistake) does not hold in this more general setting. For instance, communication does not necessarily lead to agreement and might well lead to a situation in which agents disagree and agree to disagree (i.e., disagreement is “common knowledge").

The main contribution of the paper is to define a revision rule in a multi-agent settings that applies when agents learn of something that they did not believe possible originally. This rule encompasses both cases in which agents do not face contradictions and can therefore simply refine their original beliefs as well as when they have to modify their beliefs so as to acknowledge some fact they thought impossible. The properties of the rule however do differ in these two cases : as mentioned above, disagreement might be an outcome of communication in the presence of contradictions.

2 Mutual Belief Systems : definition and preliminaries

Let $I = \{1, ..., i, ..., n\}$ be a finite set of agents and S a set of states of nature. A mutual belief system is a representation of agents' beliefs about the state of nature s and about the beliefs of the other agents. Because of this latter aspect, the structure introduced

[1]We thank Bernard Walliser for useful comments and discussion. We also thank the audience at RUD 2002 in Gif sur Yvette and participants of the workshop on Belief Revision held at LIP6, U. Paris VI. Financial support from the French Ministry of Research (ACI “Cognitique") is gratefully acknowledged.

has to be self-referential.

Definition 1 *A Mutual Belief System (MBS) 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* $2^{\Omega} \setminus \emptyset$*,*

(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).*[2]

An element $(\omega; s(\omega); t_1(\omega), ..., t_n(\omega))$ is called a *state of the world*. Its interpretation is as follows : 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 (and can also be thought of as "the type" of agent i 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 ω uniquely determines the state of the world.

The definition does *not* require that agents consider ω_0 possible, i.e., ω_0 need not be in $t_i(\omega_0)$. A consequence of allowing mistaken beliefs is that the MBS is not necessarily known by the agents. 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. Third, we assume that the mutual belief system is minimal in the sense that it does not contain a smaller MBS (condition (v)).

[2] $t_i|_{\Omega'}$ 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 $\omega \in \Omega'$.

Proposition 1 *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* $A \subset \Omega$, $t_i(A) = \cup_{\omega\in A} t_i(\omega)$.

Example 1 *Let* $S = \{\alpha, \beta\}$, $I = \{1, 2\}$ *and* $\Omega = \{\omega_0, \omega_1, \omega_2, \omega_3\}$ *such that :*

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

$$\omega_1 = (\alpha, \{\omega_1, \omega_2\}, \{\omega_1, \omega_2\})$$

$$\omega_2 = (\beta, \{\omega_1, \omega_2\}, \{\omega_1, \omega_2\})$$

$$\omega_3 = (\beta, \{\omega_3\}, \{\omega_3\})$$

This represents a situation in which the true state of nature is α, agent 1 believes that it is α or β and agent 2 believes that it is β. Furthermore, 1 believes that 2 believes that the state of nature is α or β while 2 believes that 1 believes that the state of nature is β. In a nutshell, 1 believes that it is common belief that the state of nature is α or β, while 2 believes that it is common belief that the state is β. The definition of an MBS should make it clear that the same epistemic state of the agents could be represented in various ways.

Example 2 *Let* $S = \{\alpha, \beta\}$, $I = \{1, 2\}$ *and* $\Omega = \{\omega_0, \omega_1, \omega_2, \omega_3, \omega_4\}$ *such that :*

$$\omega_0 = (\alpha, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\})$$

$$\omega_1 = (\alpha, \{\omega_1, \omega_2\}, \{\omega_1, \omega_2\})$$

$$\omega_2 = (\beta, \{\omega_1, \omega_2\}, \{\omega_1, \omega_2\})$$

$$\omega_3 = (\beta, \{\omega_3\}, \{\omega_3, \omega_4\})$$

$$\omega_4 = (\beta, \{\omega_4\}, \{\omega_3, \omega_4\})$$

An examination of the beliefs represented here reveals that the epistemic situation is the same as the one in Example 1.

A way of getting around this difficulty is to define notions of *representation* and *equivalence* of MBS as well as a notion of *minimality* for MBS. This is done in Tallon, Vergnaud, and Zamir (2003) to which we refer the interested reader. Minimality consists in essentially getting rid of potential redundancies in an MBS. Our definition thus identifies redundancies in Example 2 and suggest to "merge" states ω_3 and ω_4. A minimal MBS is one in which all the redundancies have been removed. In the rest of the paper, we exclusively deal with minimal MBS.

3 Common Belief in Mutual Beliefs Systems

In this section, we explore ways of expressing belief properties in MBS. We first define the notion of belief horizon for an agent. It is the set of states of the world that are believed possible by the agent (possibly via links of the form "I believe that you believe this state is possible", or "I believe that you believe that she believes this state is possible",...).

3.1 Belief Horizon and Common Belief

We introduce here the notion of belief horizon of an agent which is the model the agent has in mind.

Definition 2 *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MBS. The* belief horizon *of agent* $i \in I$*, denoted by* $BH_i(\omega_0, t)$*, is the minimal subset* Y *of* Ω *satisfying :*

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

Thus, $BH_i(\omega_0, t)$ is the smallest "public event" for i, i.e., the smallest set such that i believes it and believes that all other agents believe it, believe that others believe that others believe it and so forth. In Example 1, one has $BH_1(\omega_0, t) = \{\omega_1, \omega_2\}$ and $BH_2(\omega_0, t) = \{\omega_3\}$.

Proposition 2 *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MBS. For all* $i \in I$, $\forall \omega \in \Omega$, $\omega \in BH_i(\omega_0, t) \Leftrightarrow \exists r \in \mathbb{N}, \exists \{i_k\}_{k=1}^{k=r}, i_k \in I, i_r = i$ *s.th.* $\omega \in t_{i_1}(t_{i_2}(\dots(t_{i_r}(\omega_0)))$

This Proposition enables us to state a useful property of MBS, namely that an MBS is the union of agents' belief horizons and of the true state (which might not be in any agent's belief horizon).

Corollary 1 *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MBS. Then,* $\Omega = \{\omega_0\} \cup (\cup_{i\in I} BH_i(\omega_0, t))$.

The definition of common belief of an event is the usual definition of common knowledge, adapted to our setting : an event is common belief if all agents believe it, all agents believe that all agents believe it and so forth.

Definition 3 *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MBS. An event* $E \subset \Omega$ *is* common belief *(CB) if for any* $r \in \mathbb{N}$ *and any sequence* $\{i_k\}_{k=1}^{k=r}$, $i_k \in I$, $t_{i_1}(t_{i_2}(\dots(t_{i_r}(\omega_0))) \subset E$

Note that as an MBS 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 can be deduced immediately from Proposition 2.

Proposition 3 *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MBS. An event* $E \subset \Omega$ *is common belief if and only if* $BH_i(\omega_0, t) \subset E$ *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.

Corollary 2 *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be an MBS. An event* $E \subset \Omega$ *is common belief if and only if*

$$\cup_{i\in I} BH_i(\omega_0, t) \subset E \subset \Omega = \{\omega_0\} \cup (\cup_{i\in I} BH_i(\omega_0, t))$$

This corollary establishes that in an MBS, at most two events can be common belief. Ω is always commonly believed (by construction of an MBS), while $\Omega \backslash \{\omega_0\}$ 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 horizon of at least one agent.

3.2 Correct Mutual Belief Systems

A mutual belief system is correct if agents make no mistake, in the sense they all believe that the true state ω_0 is possible. Correctness, in our framework, is in some sense the analogue of the truth axiom in knowledge systems, which asserts that if an agent knows something then it must be true. However, it is possible to construct examples in which all agents are correct but this is not commonly believed. This stronger notion is captured by the notion of totally correct MBS.

Definition 4 *Let $(\Omega, \omega_0, s, (t_i)_{i\in I})$ be a minimal MBS. An agent $i \in I$ has* correct beliefs *if $\omega_0 \in t_i(\omega_0)$. The MBS is* correct *if all agents have correct beliefs. The MBS is* totally correct *if $\omega \in t_i(\omega)$ for all $\omega \in \Omega$ and all $i \in I$.*

An MBS can be correct but not totally correct, as illustrated in the following example.

Example 3 *Let $S = \{\alpha, \beta\}$ and $I = \{1, 2\}$. Consider $\Omega = \{\omega_0, \omega_1\}$ where*

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

$$\omega_1 = (\beta, \{\omega_0\}, \{\omega_0, \omega_1\})$$

In this Example, the two agents satisfy the truth axiom in the true world ω_0 but agent 2 does not believe that agent 1 satisfies it : 2 believes that in the possible world ω_1, agent 1 is mistaken. Thus there is a difference between situations where all agents satisfy the truth axiom but this fact is not commonly believed (i.e the MBS is correct but not totally correct) and situations which are captured through totally correct MBS where all agents satisfy the truth axiom and this fact is common belief.

If an agent is correct, it is easy to see that his belief horizon contains the belief horizons of all other agents, and his belief horizon is the entire space Ω. A direct corollary of this fact together with Corollary 2 is that if at least one agent is correct, the only common belief event is Ω itself. Further, since MBS that are correct have the feature that different agents' belief horizons coincide, this common belief horizon is common belief and therefore correctness embeds a kind of agreement among agents about what the model is. Note finally that correctness is sufficient but not necessary for the coincidence of belief horizons of the different agents. Take for instance the following MBS : $\omega_0 = (\alpha, \{\omega_0\}, \{\omega_1\})$ and $\omega_1 = (\beta, \{\omega_0\}, \{\omega_1\})$. There, $BH_1(\omega_0; t) = BH_2(\omega_0; t) = \{\omega_0, \omega_1\}$ while agent 2 is not correct.

4 Communication and Revision in Mutual Belief Systems

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 rules according to which agents revise their beliefs in a communication process. The analysis will concentrate on the case in which agents announce *truthfully and precisely* their beliefs.

Definition 5 *Let $(\Omega, \omega_0, s, (t_i)_{i\in I})$ be an MBS. A* communication *is a collection*

$(t_i(\omega_0))_{i\in I^c}$ where $I^c \subset I$.[3]

We'll refer to *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. Lying is hence prohibited. We will assume in the sequel that it is

[3]Thus, we restrict attention to communication that are *full truthful* in the sense that agents who communicate tell the truth, the whole truth and nothing but the truth.

"common knowledge" that agents announce precisely their true beliefs.

We now move on to some attempts to define a revision rule. We show through examples that the most intuitive rules are not adapted to our setting where agents might be mistaken.

4.1 Defining a revision procedure : a first attempt

A first intuition that one might have is simply to assume that each agent takes the announcement of the other agents at face value and hence revises his beliefs by taking the intersection of his initial beliefs with the announcement of the other. A second intuition is that an agent is not directly interested by the content of the announcements but rather by the worlds which are compatible with the announcements, i.e., he considers *the states of the world in which these announcements could have been made* ; any other state of the world is eliminated by the revision. To illustrate these two logics, consider the following example :

Example 4 *Let* $S = \{\alpha, \beta\}$ *and* $I = \{1, 2\}$ *. Consider* $\Omega = \{\omega_0, \omega_1, \omega_2\}$ *where*

$$\omega_0 = (\alpha, \{\omega_0, \omega_1\}, \{\omega_0, \omega_2\})$$

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

$$\omega_2 = (\beta, \{\omega_2\}, \{\omega_0, \omega_2\})$$

When $I^c = I$*, the two revision rules suggested above both yield* $\omega_0 = (\alpha, \{\omega_0\}, \{\omega_0\})$ *that is, both agents learn from the other's announcement that the true state is* ω_0*. However, the process through which one arrives at this MBS is different in the two rules : according to the first intuition, agent* 1 *drops state* ω_1 *because agent* 2 *announced that he does not believe in this state while according to the second intuition, agent* 1 *drops state* ω_1 *because in that state, agent* 2 *would have announced* $\{\omega_1\}$*.*

This example is actually representative of the class of totally correct MBS in which the two revision rules suggested yield the same, well-defined MBS. Before pointing differences between these two rules, we first introduce them formally.

Definition 6 *Let* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *be a totally correct MBS. Given a communication* I^c*, the revision of beliefs is the MBS,* $(\Omega^c, \omega_0, s, (t_i^c)_{i\in I})$ *defined by :*[4]

- First revision rule.
  - $\forall i \in I,\ \forall \omega \in \Omega,\ t_i^c(\omega) = t_i(\omega) \cap (\cap_{j\in I^c} t_j(\omega_0))$,
  - $\Omega^c = \{\omega_0\} \cup (\cup_{i\in I} BH_i(\omega_0, t^c))$
- Second revision rule
  - $\forall i \in I,\ \forall \omega \in \Omega,\ t_i^c(\omega) = t_i(\omega) \cap \{\omega' \in \Omega | t_j(\omega') = t_j(\omega_0);\ \forall j \in I^c\}$
  - $\Omega^c = \{\omega_0\} \cup (\cup_{i\in I} BH_i(\omega_0, t^c))$

It is readily verified that the total correctness of the MBS guarantees that the revised beliefs are well defined and hence we have :

Proposition 4 *The revision of beliefs according to the first and to the second revision rule yield, in a totally correct MBS, the same totally correct MBS.*

When we consider non totally correct MBS, we encounter two kinds of problems. First, the two rules might lead to different MBS.[5] Second, they may not be applicable. Let us examine the first problem on Example 3 where the MBS is correct.

Example 5 *(Example 3 continued)*

Let $I^c = \{1\}$*. Then the first rule leads to* $\omega_0 = (\alpha, \{\omega_0\}, \{\omega_0\})$*, while the second rule leaves the initial MBS unchanged.*

We feel that the first rule is not very satisfactory in this example. Indeed, it is as if the first agent managed to convince agent 2 to drop

[4] For convenience (or abuse...) of notation, the names of the states of the world in Ω^c are the same as in Ω (but with different beliefs of course.)

[5] Although we defined the two rules only for totally correct MBS, it is clear that they are applicable to a wider set of MBS.

state ω_1 : 1 says "I believe the state of nature is α" and agent 2 is convinced that he should not consider any more that the state of nature could be β. Yet, before any communication took place, agent 2 thought that agent 1 could well be mistaken on the state of nature and agent 1's announcement was completely foreseeable by agent 2. Thus, following the first rule leads to admit that agent 2 is influenced by others' announcement even though it was expected and hence is not confident in his own beliefs. The second revision rule leaves the MBS unchanged which looks more reasonable.

In view of this example, we chose to generalize the second intuition rather than the first. This amounts to implicitly assume that one will never abandon one's initial beliefs when they are not proven false. Even if an agent's beliefs are contradicted by the beliefs of another agent, the first agent will not adopt the second agent's beliefs but simply incorporate in his own beliefs the fact that they disagree. There is a sense in which revised beliefs are entrenched in the initial beliefs. This represents situations in which each agent believes that his own expertise is at least as good as the others'.

The second problem, which is faced by the two rules, is that they might be ill-defined for non correct MBS, as shown in the following example.

Example 6 *Let* $S = \{\alpha, \beta, \gamma\}$ *and* $I = \{1, 2\}$ *. Consider* $\Omega = \{\omega_0, \omega_1\}$

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

$$\omega_1 = (\beta, \{\omega_1\}, \{\omega_1\})$$

Assume full communication. Then, following the first rule the revision yields $t_1^c(\omega_0) = t_2^c(\omega_0) = \emptyset$ *and following the second rule, the revision yields* $t_2^c(\omega_0) = \{\omega_0\}$ *while* $t_1^c(\omega_0) = \emptyset$*, which is not possible in an MBS.*

The problem with the second rule exhibited in Example 6 reflects the contradiction between agent 1's initial beliefs and his interpretation of the other agent's announcement. Observe that the first revision rule has the same problem and cannot be applied here either. We reached now the difficult part in the construction of a general revision rule namely, the need to specify how agents *deal with contradictions* between their initial beliefs and the reported beliefs of the other agents.

4.2 Coping with contradictions

The two revision rules introduced above are formally not applicable when contradictions occur, that is, if there is no world among the ones initially believed by an agent that is compatible with the announcements of the other agents. However, the logic behind these two rules could be extended to deal with contradictions. Along the intuition of the first revision rule, the agent could adopt the beliefs announced by the other. In Example 6, agent 1's beliefs would now be given by $t_1^c(\omega_0) = \{\omega_0\}$. This corresponds to take at face value 2's announcements and, in particular, to admit that state α is true, something 1 did not believe in to begin with. Observe however that agent 1 is not proven wrong in his belief that the state of nature is β. Indeed, the only mistake that is revealed is that 1 believed 2 believed the state of nature was α. We do not pursue this logic in the rest of the paper and concentrate on the logic behind the second revision rule : in presence of contradictions, the agent holds on to the beliefs that are not contradicted and changes in a "minimal way" the ones that are contradicted. Consider Example 6 again : what does agent 2 do when he hears 1's announcement that the state is α, which is contradicting his initial beliefs ? A plausible revised MBS, after full communication is :

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

$$\omega_1 = (\beta, \{\omega_1\}, \{\omega_0\})$$

This revision is minimal in the sense that the initial disagreement on the state of nature persists. Agent 2 has only revised his beliefs by taking into account 1's beliefs, which explains why $t_2^c(\omega_1) = \{\omega_0\}$. The system is then closed by imposing that this minimal change becomes common belief. The general revision rule we'll introduce shortly is hence built on the idea that an agent keeps the beliefs that are not contradicted. This is illustrated in the next example.

Example 7 *Let* $S = \{\alpha, \beta, \gamma\}$ *and* $I = \{1, 2, 3\}$. *Consider* $\Omega = \{\omega_0, \omega_1, \omega_2\}$

$$\omega_0 = (\alpha, \{\omega_1\}, \{\omega_2\}, \{\omega_0\})$$

$$\omega_1 = (\beta, \{\omega_1\}, \{\omega_1\}, \{\omega_1\})$$

$$\omega_2 = (\gamma, \{\omega_1\}, \{\omega_2\}, \{\omega_2\})$$

Assume communication $I^c = \{2\}$. *Agent 1 realizes that he was mistaken about 2's beliefs : 2 actually disagrees with 1 on both the state of nature and 3's beliefs. A plausible revision would be :*

$$\omega_0 = (\alpha, \{\omega_1\}, \{\omega_2\}, \{\omega_0\})$$

$$\omega_1 = (\beta, \{\omega_1\}, \{\omega_2\}, \{\omega_1\})$$

$$\omega_2 = (\gamma, \{\omega_1\}, \{\omega_2\}, \{\omega_2\})$$

where 1 only modified his beliefs about 2's beliefs (and the latter are common belief).

We are not yet done : the principles discussed so far are still not enough to yield a satisfactory revision rule, as can be seen on the following example.

Example 8 *Let* $S = \{\alpha, \beta, \gamma\}$ *and* $I = \{1, 2, 3\}$. *Consider* $\Omega = \{\omega_0, \omega_1, \omega_2\}$

$$\omega_0 = (\alpha, \{\omega_1, \omega_2\}, \{\omega_0\}, \{\omega_1\})$$

$$\omega_1 = (\beta, \{\omega_1, \omega_2\}, \{\omega_1, \omega_2\}, \{\omega_1\})$$

$$\omega_2 = (\gamma, \{\omega_1, \omega_2\}, \{\omega_1, \omega_2\}, \{\omega_2\})$$

Assume communication $I^c = \{2, 3\}$. *Agent 1 realizes he was mistaken about 2's beliefs but not necessarily about 3's beliefs. Thus, he will not necessarily keep all his initial beliefs* $\{\omega_1, \omega_2\}$*, and for instance, he might abandon* ω_2*, yielding the following revised MBS :*

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

$$\omega_1 = (\beta, \{\omega_1\}, \{\omega_0\}, \{\omega_1\})$$

To capture the phenomenon at work in Example 8, we need to add a sort of *personal attitude* of the agents as part of the data of the model. To capture this personal attitude of the agent, we assume that given an MBS, $(\Omega, \omega_0, s, (t_i)_{i\in I})$ and a communication I^c, there is an *order*, $\succeq_i^c$ for each i, which is a complete and transitive binary relation defined on Ω_i, the set of states that i could believe (recall that $\Omega_i = \cup_{\omega\in\Omega} t_i(\omega)$). For $\omega, \omega' \in \Omega_i$, $\omega \succeq_i^c \omega'$ is interpreted as saying that, given the communication I^c, agent i believes that ω is "closer" to the true (unknown) world than ω' is. In addition to completeness and transitivity we shall assume that the order $\succeq_i^c$ is *consistent* :

Definition 7 *Given an MBS,* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *and a communication* I^c*, an order* $\succeq_i^c$ *is said to be* consistent *if whenever* $\Omega_i \cap (\{\omega \in \Omega | t_j(\omega) = t_j(\omega_0), \forall j \in I^c\}) \neq \emptyset$, $[\omega \in \Omega_i \cap \{\omega'' \in \Omega | t_j(\omega'') = t_j(\omega_0), \forall j \in I^c\}]$

$$\Leftrightarrow \omega \succeq_i^c \omega';\ \ \forall \omega' \in \Omega_i w.$$

Thus, an order $\succeq_i^c$ is consistent if it ranks highest every state of the world that is initially deemed to be possibly believed by i (i.e., states that are in Ω_i) and that explains (is compatible with) the others' announcements. This requirement is really rather weak : in Example 8, it does not impose anything on $\succeq_1^c$ and the three different orders $\omega_1 \succ_1^c \omega_2$, $\omega_2 \succ_1^c \omega_1$, or $\omega_1 \sim_1^c \omega_2$ are all consistent.

The revision rule we are about to introduce is based on these orders and on the assumption that, loosely speaking, they are commonly believed by all agents so as to enable interactive reasoning about mutual beliefs. It should be noted that i's order is defined on Ω_i, which does not, in general, coincide with i's belief horizon. This is important since j might (mistakenly) believe that there are states that i considers possible (i.e., states in $BH_j(\omega_0, t) \cap \Omega_i$). Hence, j needs to know how to revise i's beliefs in these worlds. Implicit in the fact that the order introduced for agent i is defined on all of Ω_i is the idea that all agents agree on how to revise i's beliefs.

If states ω and ω' belong to both j and j''s belief horizon, then these two agents agree on which is ranked highest according to i's ordering. Furthermore, this fact is commonly believed by all agents. Hence, we'll make the maintained assumption that given an MBS, $\left(\Omega, \omega_0, s, (t_i)_{i\in I}\right)$ and a communication, all orders $\succeq_i^c$ for $i \in I$ are consistent and commonly known by all agents (in the sense we just discussed).

4.3 A general revision rule : definition and examples

We now propose a general revision rule that copes with announcements contradicting initial beliefs. We first define the rule and then illustrate it via a few examples. Given an MBS and a communication, the revision rule we propose consists of two elements.

- **Step 1** Each agent i retains all states of the world in the set Ω_i that have the highest rank in his order $\succeq_i^c$.
- **Step 2** In the states retained, the beliefs attributed to other agents are constructed by taking into account the modifications they have made in step 1. This corresponds to the idea that the way agents modify their beliefs is common belief.

Note that step 2 is possible since by assumption the announcements $t_i(\omega_0)$, $i \in I^c$, and the orders $\succeq_i^c$ are " commonly known", hence the modification made can be "performed" by player i for each player j.

Definition 8 *Let $\left(\Omega, \omega_0, s, (t_i)_{i\in I}\right)$ be a minimal MBS and consider the communication I^c. The* revision *of $\left(\Omega, \omega_0, s, (t_i)_{i\in I}\right)$ is $\left(\Omega^c, \omega_0, s, (t_i^c)_{i\in I}\right)$, defined in two steps :*

First define $\tilde{t}_i(.)$ by $\forall \omega \in \Omega$, $\tilde{t}_i(\omega) = \{\omega' \in t_i(\omega) | \omega' \succeq_i^c \omega'', \forall \omega'' \in t_i(\omega)\}$ and let $\widetilde{\Omega} = \{\omega_0\} \cup \left(\cup_{i\in I} BH_i(\omega_0, \tilde{t})\right)$

Then, define $t_i^c(.)$ as follows :

- $\forall \omega \in \Omega$, $\forall i \in I \setminus I^c$, $t_i^c(\omega) = \tilde{t}_i(\omega)$
- $\forall \omega \in \Omega$, $\forall i \in I^c$, $t_i^c(\omega) = \tilde{t}_i(\omega_0)$

and set $\Omega^c = \{\omega_0\} \cup (\cup_{i\in I} BH_i(\omega_0, t^c))$

The logic of the revision rule we propose can be understood as follows. The first step is to eliminate in one's beliefs the worlds that are considered the farthest away (after hearing the others' announcements) from the true world, according to the order $\succeq_i^c$. In this operation, the agent considers his initial beliefs as valid and simply gets rid of the states that are not ranked highest with respect to his ordering. Hence, as discussed in section 4.2, agents are assumed to anchor their revised beliefs in their initial beliefs. This step of the revision procedure coincides with the second revision rule when it is well defined and the agents' orders are consistent.

The second step of the revision procedure consists in dealing with the remaining "contradictions". After the first step of the revision procedure, agents announce their (corrected) beliefs : if, after the first step, the MBS obtained contains states that specify different beliefs for say agent i than the ones he announces, then simply replace these beliefs by his announcement. This second step might be irrelevant if the announcements of the agents were all compatible with what they expected (this is the case under our maintained assumption for instance if the MBS is correct), in which case after the first step, all the states that were not ranked highest have been eliminated. We illustrate this on Example 1.

Example 9 *(Example 1 continued) Consider full communication and the following consistent orders for 1 and 2 :*

$$\omega_3 \succ_1^c \omega_1 \succ_1^c \omega_2 \text{ and } \omega_1 \sim_2^c \omega_2 \succ_2^c \omega_3$$

The first step of the definition yields the following MBS :

$\omega_0 = (\alpha, \{\omega_1\}, \{\omega_3\})$

$\omega_1 = (\alpha, \{\omega_1\}, \{\omega_1, \omega_2\})$

$\omega_2 = (\beta, \{\omega_1\}, \{\omega_1, \omega_2\})$

$\omega_3 = (\beta, \{\omega_3\}, \{\omega_3\})$

At the next step, the contradictions are treated by replacing with the announcement.

$\omega_0 = (\alpha, \{\omega_1\}, \{\omega_3\})$

$\omega_1 = (\alpha, \{\omega_1\}, \{\omega_3\})$

$\omega_2 = (\beta, \{\omega_1\}, \{\omega_3\})$

$\omega_3 = (\beta, \{\omega_1\}, \{\omega_3\})$

It is easy to check that in the above MBS, states ω_0 and ω_1 actually express the same hierarchy of beliefs, and that the same is true for states ω_2 and ω_3. Hence, it can be reduced (according to the formal process defined in Appendix A) to the following MBS :

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

$\omega_1 = (\beta, \{\omega_0\}, \{\omega_1\})$

The outcome of the revision procedure is therefore a situation in which disagreement about the state of nature is common belief but becomes common belief.

4.4 General revision rule : consistency properties

Contrary to the rules discussed in Section 4.1, revision is always possible and does not lead to ill-defined belief systems.

Proposition 5 *Let $\left(\Omega, \omega_0, s, (t_i)_{i\in I}\right)$ be a minimal MBS. Then, $\left(\widetilde{\Omega}, \omega_0, s, \left(\tilde{t}_i\right)_{i\in I}\right)$ and $\left(\Omega^c, \omega_0, s, (t_i^c)_{i\in I}\right)$ are MBS.*

The next Proposition establishes a link between the second revision rule and the general one. This formally shows that the logic behind the general revision rule is indeed, as argued above, the one present in the second revision rule.

Proposition 6 *Assume that agents' orders are consistent. Then, when the second revision rule is applicable, it coincides with the first step of the general revision rule, while the second step is void.*

A direct corollary to this Proposition is that when the MBS is totally correct then the second step of the revision process is void (i.e., $\widetilde{\Omega} = \Omega^c$) if the agents' orderings over the state space are consistent. This proposition also establishes that when the MBS is totally correct (a sufficient condition for the first two rules introduced above to be well defined), then all the revision rules we have defined coincide.

4.5 General revision rule : agreement and consensus

In this section, we seek to characterize conditions under which the revision leads to different forms of agreements among agents. This requires making a detour *via* the definition and characterization of *common S-beliefs systems*, in which agents' beliefs about the state of nature are common belief.

For a given MBS, $(\Omega, \omega_0, s, (t_i)_{i\in I})$ define the S-belief to be the event

$SB(\omega_0, t) =$

$\{\omega \in \Omega | s(t_i(\omega)) = s(t_i(\omega_0)) \; \forall i \in I\}$

The S-belief is the event "for all $i \in I$, agent i believes that the *state of nature* is in $s(t_i(\omega_0))$". In other words, $SB(\omega_0, t)$ is the subset of Ω in which the first level beliefs about S are as those in ω_0, i.e., the beliefs in the true state. We define now a special case of belief system, where the first level beliefs about S are common beliefs.

Definition 9 *An MBS, $(\Omega, \omega_0, s, (t_i)_{i\in I})$ is a* common S-belief system *(CSBS) if $SB(\omega_0, t)$ is common belief.*

In a CSBS, the agents' beliefs about the state of nature are common beliefs. Agents need not agree in a CSBS. It is thus possible to represent situations in which agents' disagreement is common belief. Example 3 is an instance of such a situation : 1 believes α, 2 believes α or β and this is common belief, i.e., agents disagree and this disagreement is common belief. We now establish properties about the degree to which agents agree after communication and revision have occurred. When *all* agents communicate, the revision leads to a situation in which beliefs about the state of nature are common belief. When agents still disagree about the state of nature, this models situation in which this disagreement is common belief.

Proposition 7 *Let* $\left(\Omega, \omega_0, s, (t_i)_{i\in I}\right)$ *be a minimal MBS. Then,* $\left(\Omega^c, \omega_0, s, (t_i^c)_{i\in I}\right)$ *is a CSBS whenever* $I^c = I$.

When the initial MBS is already a CSBS, that is, when the beliefs about the state of nature of all agents are common belief, then communication does not lead to any further revision.

Proposition 8 *Let* $\left(\Omega, \omega_0, s, (t_i)_{i\in I}\right)$ *be a minimal CSBS. Then, if agents' orderings* $\succeq_i^c$ *are consistent,* $\left(\Omega^c, \omega_0, s, (t_i^c)_{i\in I}\right) = \left(\Omega, \omega_0, s, (t_i)_{i\in I}\right)$

The notion of CSBS does not entail a strong notion of agreement since indeed, disagreement can be common belief. A particular case of a CSBS is when the first level beliefs of all agents are the same, and thus $t_i(\omega) = t_j(\omega)$ for all $i, j \in I$ and all $\omega \in \Omega$. This represents a situation of consensus, when all agents have the same beliefs.

Definition 10 *A minimal MBS,* $(\Omega, \omega_0, s, (t_i)_{i\in I})$ *is* consensual *if for all* $i, j \in I$, $t_i(\omega_0) = t_j(\omega_0)$.[6]

We now give a sufficient condition that entails that revision leads to a consensual MBS.

Proposition 9 *Let* $\left(\Omega, \omega_0, s, (t_i)_{i\in I}\right)$ *be a minimal MBS and assume it is totally correct. Assume further that* $I^c = I$ *and that agents' orderings* $\succeq_i^c$ *are consistent. Then,* $\left(\Omega^c, \omega_0, s, (t_i^c)_{i\in I}\right)$ *is consensual.*

[6]If the MBS is not minimal then an MBS is said to be consensual if it has a representation that is consensual.

This Proposition establishes that only under rather strong assumption will the revision process lead to a consensual belief system, in which all agents agree. Indeed, the assumption that the MBS be totally correct is necessary to get consensus, as can be seen on Example 3, in which the MBS is correct but not totally correct and no revision occurs after full communication.

Références

ALCHOURRON, C., P. GARDENFORS, AND D. MAKINSON (1985) : "On the logic of theory change : partial meet contraction and revision functions," *The Journal of Symbolic Logic*, 50, 510–530.

AUMANN, R. (1976) : "Agreeing to disagree," *Annals of Statistics*, 4(6), 1236–1239.

GEANAKOPLOS, J. (1994) : "Common knowledge," in *Handbook of game theory*, ed. by R. Aumann, and S. Hart, vol. 2, chap. 40, pp. 1437–1496. North Holland - Amsterdam.

GEANAKOPLOS, J., AND H. POLEMARCHAKIS (1982) : "We can't disagree forever," *Journal of Economic Theory*, 28, 192–200.

MERTENS, J.-F., AND S. ZAMIR (1985) : "Formulation of Bayesian analysis for games with incomplete information," *International Journal of Game Theory*, 10, 619–632.

TALLON, J.-M., J.-C. VERGNAUD, AND S. ZAMIR (2003) : "Contradicting beliefs and communication," Working paper 2003-13, EUREQua-Université Paris I, ftp ://mse.univ-paris1.fr/pub/mse/cahiers2003/V03013.pdf.