Mutual Belief Revision: Semantics and Computation

Yi Jin
School of Computing Science
Simon Fraser University, Canada
yij@cs.sfu.ca

Michael Thielscher
Department of Computer Science
Technische Universität Dresden, Germany
mit@inf.tu-dresden.de

Dongmo Zhang
School of Computing and Mathematics
University of Western Sydney, Australia
dongmo@scm.uws.edu.au

Abstract

This paper presents both a semantic and a computational model for multi-agent belief revision. We show that these two models are equivalent but serve different purposes. The semantic model displays the intuition and construction of the belief revision operation in multi-agent environments, especially in case of just two agents. The logical properties of this model provide strong justifications for it. The computational model enables us to reassess the operation from a computational perspective. A complexity analysis reveals that belief revision between two agents is computationally no more demanding than single agent belief revision.

Introduction

It is an interesting problem how people understand each other through information exchange. Bypassing communication and psychology issues, the problem can be described as a pure AI issue as "how epistemic agents in a multi-agent system revise their beliefs as a result of belief exchange". The classical study of belief revision has been focused on how a single agent revises her beliefs to incorporate new information. This research normally assumes that new information is fully accepted, no matter whether it is represented as a single formula or as a set of sentences. Obviously such an assumption is not applicable to multi-agent systems. There have been a variety of approaches proposed in the literature to deal with the problem of belief revision in multi-agent systems. The approach of *non-prioritized belief revision* allows an agent to revise her beliefs by partially accepting new information (Hansson 1999). This research sheds light on belief change with defeasible information resources but still focuses on the formalization of belief change from a single agent perspective. The acceptance or rejection of information resources is purely determined by the epistemic subject rather than decided by all participating agents. The study of *belief merging* or *knowledge arbitration* directly accounts for multi-agent belief change by pursuing a "fair" process that is able to incorporate individual beliefs of agents into coherent group beliefs (Revesz 1997; Liberatore & Schaerf 1998; Konieczny & Pérez 1998). These approaches sometimes force agents to accept some democracy rules, such as majority, equal reliability, social welfare maximization and so on, but disregard an agent's "personal" view of group consentience. The present paper takes a different perspective to deal with the problem of multi-agent belief revision. We consider belief change in a multi-agent system a two-stage process. In the first stage, all agents sit together to work out a mutually acceptable point of view (a common understanding of the world) through a sequence of interchanges of respective views. This process is similar to belief merging. Once such a common understanding has been reached, in the second stage of belief change each agent will adjust her original belief state in order to form a new view of the world as a result of belief interchange. What distinguishes our approach from belief merging is that the agents do not always reach a consensus after mutual belief revision, although their beliefs will in general be "closer" to each other. This view is indirectly supported by the observation that we, human beings, do not always completely agree with each other after exchanging our opinions on certain issues. Note that as the participating agents will have "closer" views of the outside world, mutual belief revision facilitates the potential cooperation between these agents. Another advantage of performing mutual belief revision is that heterogeneous agents can share their different sensing capabilities, which of course helps to maximize the overall utility of a multi-agent system.

To make our exploration simple, we will focus on the belief revision problem in the setting of two-agent systems, so-called *mutual belief revision*. We will present a semantic model based on Ordinal Conditional Function (OCF) (Spohn 1988) to specify the above-mentioned two stages of mutual belief revision. While the OCF model is conceptually clear and constructively simple, it is not computation-friendly because it requires to take an exponential input of possible worlds from each agent. In order to have an estimation of the computational complexity, we will present another model of mutual belief revision based on a sort of generalized epistemic entrenchment ordering (Gärdenfors & Makinson 1988). We will prove that this model is essentially equivalent to the OCF model, and we will give first complexity results.

The plan of the paper is as follows. In next section, we present an OCF-based semantic model for mutual belief

revision operation, followed by a discussion of its formal properties. Then we will present a computational model for mutual belief revision and some results on its computational complexity. We conclude the paper with a brief discussion of related work.

We will express our theory of mutual belief revision in terms of a propositional language $\mathcal{L}$. The language is that of classical propositional logic with an associated consequence operation *Cn* in the sense that $Cn(X) = \{\alpha \mid X \vdash \alpha\}$, where X is a set of sentences. A set K of sentences is *logically closed* or called a *belief set* whenever $K = Cn(K)$. If X, Y are two sets of sentences, $X+Y$ denotes $Cn(X \cup Y)$. The set of all propositional interpretations (possible worlds) of $\mathcal{L}$ is denoted by $\mathcal{W}$. A sentence α is true in a world w, written as $w \models \alpha$, iff w makes α true in the classical, truth functional way. A set X of sentences is true in a world $w \in \mathcal{W}$, denoted by $w \models X$, iff every element in X is true in w. Finally, the i-th projection of a function M is denoted by M_i, that is, $M_i(\overline{x}) = P_i(M(\overline{x}))$ where P_i is the i-th projection function.

OCF Model of Mutual Belief Revision

In this section, we will explain the concept of mutual belief revision and present a formal, semantic model for this concept by using Spohn's *Ordinal Conditional Functions* (OCF), which have been widely employed in the literature (Spohn 1988).

Belief States in OCF Model

With Spohn's original model of OCF a belief state of an agent is represented as a function k which maps the set of all possible worlds to a class of ordinals. For the sake of simplicity, we consider here so-called *Natural Conditional Functions* (Spohn 1991), in which the range of an OCF k is the set of natural numbers, that is, $k : \mathcal{W} \to \mathbb{N}$. The set of all OCFs is denoted by κ.

Intuitively, an OCF represents the degree of plausibility of possible worlds, or more precisely, the grading of disbelief. The lower the number, the more plausible is a world. Representing beliefs by means of an OCF provides a richer structure than the set representation of beliefs, in the sense that it encodes both a belief set and the plausibility of beliefs. This is not only useful for modeling iterated belief revision (Darwiche & Pearl 1997) but also, as we will see, for modeling mutual belief revision.

Given an OCF k and a natural number i, the set of worlds with ranks smaller or equal to i is called a *sphere* of i with radius i, denoted by $k^-(i)$. Formally,

$$k^-(i) = \{w \mid k(w) \leq i\}$$

In particular, $k^-(0)$ is called the *core* of k. As usual, the belief set of a belief state k, denoted by $Bel(k)$, is the set of sentences which hold in the core of k:

$$Bel(k) = \{\alpha \in \mathcal{L} \mid w \models \alpha \text{ for all } w \in k^-(0)\}$$

Therefore, we say that two OCFs k_1, k_2 are epistemically equivalent, denoted by $k_1 \equiv k_2$, iff $k_1^-(0) = k_2^-(0)$. Two OCFs k_1, k_2 are consistent iff $k_1^-(0) \cap k_2^-(0) \neq \emptyset$.

An OCF belief state also encodes an epistemic entrenchment (EE) ordering: For any sentence β,

$$Rank_k(\beta) = \begin{cases} \infty, & \text{if } \vdash \beta; \\ \min\{k(w) \mid w \models \neg\beta\}, & \text{otherwise.} \end{cases} \quad (1)$$

It is easy to see that $Bel(k) = \{\beta \mid Rank_k(\beta) > 0\}$. Moreover, if we define an ordering $\leq$ over the language by "$\alpha \leq \beta$ iff $Rank_k(\alpha) \leq Rank_k(\beta)$", then $\leq$ satisfies Postulates (EE1)-(EE5) of (Gärdenfors & Makinson 1988).

Reaching a Common Understanding

We consider mutual belief revision to be a two-stage process. In the first stage, two agents try to reach a common understanding (which is assumed to be logically consistent) through sequential belief interchange. Once such a common understanding is reached, each agent performs a belief revision process to adapt his or her belief state to the information learnt from the other agent. We will model these two stages separately.

The first stage of mutual belief revision is very similar to *belief merging*. To reach a consistent common understanding, several rounds of "belief interchange" might be needed. In each round, each agent receives more information from the other one so that both get a better and better "understanding" of each other.

Given two pairs of subsets of possible worlds $\langle s, t\rangle$ and $\langle s', t'\rangle$, we say that $\langle s', t'\rangle$ is *closer* than $\langle s, t\rangle$, denoted by $\langle s, t\rangle < \langle s', t'\rangle$, iff $s \subseteq s'$, $t \subseteq t'$ and $s \cup t \subset s' \cup t'$.

Consider two belief states k_1 and k_2 of two agents. We start with the cores of k_1 and k_2; if they do not intersect, the next round of belief interchange will continue on the least radius r such that $\langle k_1^-(0), k_2^-(0)\rangle < \langle k_1^-(r), k_2^-(r)\rangle$, requiring to check the intersection of the corresponding two spheres $k_1^-(r)$ and $k_2^-(r)$. The process continues until a radius is reached which is large enough for the corresponding spheres to intersect.[1] The procedure can be represented by the *sequence of belief interchange* $\langle k_1^-(0), k_2^-(0)\rangle, \ldots, \langle k_1^-(r_n), k_2^-(r_n)\rangle$, where

1. $r_{i+1} = \min\{r \mid \langle k_1^-(r_i), k_2^-(r_i)\rangle < \langle k_1^-(r), k_2^-(r)\rangle\}$;
2. $r_n = \min\{r \mid k_1^-(r) \cap k_2^-(r) \neq \emptyset\}$.

This sequence clearly shows the procedure of "mutual understanding": each agent gradually broadens their views (possible worlds) in order to reach a common understanding (intersection of possible worlds from two agents). Afterwards, the belief state of each agent with the common understanding can be represented, respectively, by belief states $k_1 - r_n$ and $k_2 - r_n$, where

$$(k - r)(w) = \begin{cases} 0, & \text{if } k(w) \leq r; \\ k(w), & \text{otherwise.} \end{cases}$$

Therefore, the first stage of mutual belief revision defines a function γ which takes a pair of belief states (possibly inconsistent with each other) and returns a pair of consistent

[1] Note that it is always possible to reach an intersection in a finite number of steps since the natural numbers are well-ordered. In the special case that $k_1^-(0)$ and $k_2^-(0)$ intersect, the process terminates immediately.

belief states:

$$\gamma(k_1, k_2) = \langle k_1 - r_n, k_2 - r_n \rangle \tag{2}$$

where $r_n = \min\{r \mid k_1^-(r) \cap k_2^-(r) \neq \emptyset\}$.

Revision of OCFs

As we described in the previous section, the first stage of mutual belief revision results in a pair of weakened belief states, which represents the (mutually consistent) remaining beliefs of the two agents. In the second stage of mutual belief revision, the weakened belief state of one agent will be sent to the other agent as the new information, and vice versa (Zhang *et al.* 2004). Both agents will review the new information they just received in the light of their original beliefs, to form a new view of the world. This process is similar to the standard operation of single-agent belief revision. The only difference is that while an agent revises her beliefs, she not only tries to incorporate the information received from the other agent but also takes the other agent's view of the information into account. In other words, the agent views the new information as a belief state rather than a single sentence or a set of beliefs. In order to model such a process we will define a belief revision operator which allows to revise an OCF by another OCF.

Let us first recall the idea of how belief revision is generated by an OCF. Following (Spohn 1988), a new evidence α is assumed to come along with an evidence degree $m > 0$. An OCF k is then revised according to new evidence α as follows:

$$(k^*_{\alpha,m})(w) = \begin{cases} k(w) + m, & \text{if } w \models \neg\alpha; \\ k(w) - \mathit{Rank}_k(\neg\alpha), & \text{otherwise}. \end{cases} \tag{3}$$

As has been shown by (Jin & Thielscher 2007), this revision operation satisfies all AGM postulates as well as the DP postulates (Darwiche & Pearl 1997) and the *independence* postulate (Jin & Thielscher 2007) for iterated revision.

Now we define a belief revision operator which allows us to revise a belief state by another belief state.

Definition 1. *Given two OCFs k and λ, we define the revision of k by λ, denoted by $k \otimes \lambda$, as follows:*

$$(k \otimes \lambda)(w) = \begin{cases} k(w) + \lambda(w), & \textit{if } \lambda(w) > 0; \\ k(w) - m_{k,\lambda}, & \textit{otherwise}. \end{cases} \tag{4}$$

where $m_{k,\lambda}$ is the smallest radius of a sphere of k which intersects the core of λ:

$$m_{k,\lambda} = \min\{i \mid k^-(i) \cap \lambda^-(0) \neq \emptyset\} \tag{5}$$

The idea behind this definition is the following. For any world which the other agent disbelieves, the agent will degrade the world according to the other agent's degree of disbelief. Conversely, for those worlds which the other agent believes, the agent will upgrade these worlds according to the original degree of belief.

We say that an OCF k encodes a sentence α with evidence degree $m > 0$ iff for any possible world w,

$$k(w) = \begin{cases} 0, & \text{if } w \models \alpha; \\ m, & \text{otherwise}. \end{cases} \tag{6}$$

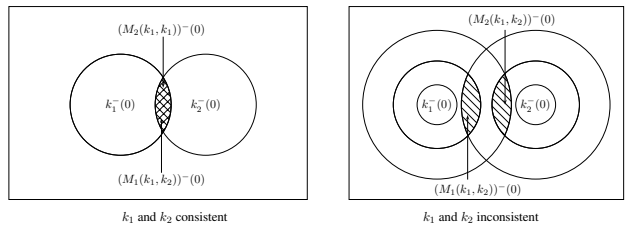

Figure 1: Visualization of mutual belief revision

It is easy to see that an OCF k encodes α with plausibility m iff $\mathit{Bel}(k) = \mathit{Cn}(\{\alpha\})$ and $\mathit{Rank}_k(\alpha) = m$. It is not difficult to see that the revision operator defined by (4) is indeed a generalization of that defined by (3).

Proposition 1. *Let k be an arbitrary OCF and λ an OCF that encodes $\langle \alpha, m \rangle$. Then for any possible w,*

$$(k^*_{\alpha,m})(w) = (k \otimes \lambda)(w)$$

OCF Model of Mutual Belief Revision

In the second stage of mutual belief revision, each agent revises her belief state by mutually accepting the point of view reached in the first stage. The construction is similar to the approach of (Zhang *et al.* 2004).

Definition 2. *A function $M : \kappa \times \kappa \to \kappa \times \kappa$ is an OCF* mutual belief revision *if for any k_1, k_2,*

$$M(k_1, k_2) = \langle k_1 \otimes \gamma_2(k_1, k_2), k_2 \otimes \gamma_1(k_1, k_2) \rangle \tag{7}$$

where γ is defined by (2).

The two diagrams in Figure 1 illustrate how the construction works. The left diagram shows a special case where k_1 and k_2 are mutually consistent. It is obvious that the first stage ends after just one round: $\pi = (\langle k_1^-(0), k_2^-(0) \rangle)$. Therefore, $M(k_1, k_2) = \langle k_1 \otimes k_2, k_2 \otimes k_1 \rangle$. The cores of both new OCFs, i.e., $M_1(k_1, k_2)$ and $M_2(k_1, k_2)$, are identified with the intersection of $k_1^-(0)$ and $k_2^-(0)$. The diagram on the right-hand side illustrates the more general situation where the initial belief states k_1 and k_2 are mutually inconsistent. For this example we assume that the belief interchange requires a total of three rounds, that is, $\pi = (\langle k_1^-(0), k_2^-(0) \rangle, \langle k_1^-(r_1), k_2^-(r_1) \rangle, \langle k_1^-(r_2), k_2^-(r_2) \rangle)$. Then $M(k_1, k_2) = \langle k_1 \otimes (k_2 - r_2), k_2 \otimes (k_1 - r_2) \rangle$. The core of $M_1(k_1, k_2)$ is exactly the intersection of $k_2^-(r_2)$ with the innermost sphere of k_1 (in our example, $k_1^-(r_1)$) which intersects $k_2^-(r_2)$.[2] Analogously, the core $M_2(k_1, k_2)$ is the intersection of $k_1^-(r_2)$ with the innermost sphere of k_2 which intersects $k_1^-(r_2)$. It is easy to see that the intersection $k_1^-(r_2) \cap k_2^-(r_n)$ contains both the core of $M_1(k_1, k_2)$ and that of $M_2(k_1, k_2)$; in other words, the common understanding is accepted by both agents after mutual belief revision.

[2]The reader may observe the analogy to the belief revision model based on Systems of Spheres (Grove 1988).

Properties of Mutual Belief Revision

To justify the OCF model, in this section we discuss some of its formal properties.

Given two OCFs k_1, k_2, we define their *inconsistency degree*, $d_{inc}(k_1, k_2)$, as follows:

$$d_{inc}(k_1, k_2) = \min\{n \mid k_1^-(n) \cap k_2^-(n) \neq \emptyset\}$$

The following result shows that belief states of agents become "more consistent" after mutual belief revision unless both agents stick firmly to their original beliefs.

Proposition 2. *The OCF mutual belief revision given by Definition 2 satisfies the following properties:*

1. $d_{inc}(M(k_1, k_2)) \leq d_{inc}(k_1, k_2)$;
2. $d_{inc}(M(k_1, k_2)) = d_{inc}(k_1, k_2)$ *just in case* $Bel(k_1) \subseteq Bel(M_1(k_1, k_2))$ *and* $Bel(k_2) \subseteq Bel(M_2(k_1, k_2))$.

The above properties of mutual belief revision are very promising. Note that as the participating agents will have "closer" views of the outside world, mutual belief revision facilitates the potential cooperation between these agents.

The simple construction of the OCF model allows us to prove more interesting properties:

Proposition 3. *The OCF mutual belief revision given by Definition 2 satisfies the following properties (for each* $i \in \{1, 2\}$*):*

(M1) $Bel(\gamma_1(k_1, k_2)) + Bel(\gamma_2(k_1, k_2)) \subseteq Bel(M_i(k_1, k_2))$;

(M2) *if* k_1 *and* k_2 *are consistent, then* $Bel(M_i(k_1, k_2)) = Bel(k_1) + Bel(k_2)$;

(M3) k_i *and* $M_i(k_1, k_2)$ *are consistent iff* $Bel(k_i) \subseteq Bel(M_i(k_1, k_2))$;

(M4) $M_1(k_1, k_2) = M_1(k_1, \gamma_2(k_1, k_2))$, $M_2(k_1, k_2) = M_2(\gamma_1(k_1, k_2), k_2)$;

(M5) *if both* $Bel(k_1) \subseteq Bel(M_1(k_1, k_2))$ *and* $Bel(k_2) \subseteq Bel(M_2(k_1, k_2))$, *then* $M_i(k_1, k_2) \equiv M_i(M(k_1, k_2))$.

(M1) ensures that the common understanding is accepted by both agents. (M2) captures the cooperative attitude of agents: if two agents have no disagreement, then each of them will accept the beliefs of the other agent. (M3), on the other hand, captures the self-interest of agents, that is, if an agent is not going to accept any beliefs that contradict her own, she does not need to give up any of her beliefs.

(M4) shows that the information an agent gains from mutual belief revision is no more than what she agrees on. In fact, (M1) and (M4) are two principal properties of mutual belief revision: both agents benefit from mutual belief revision without loss of the diversity of views. In general, it is not necessary that the two belief states will merge; as described by (M5), the agents may get stuck in a stand-off (fixed-point) if none of them is willing to make concessions.

Computational Mutual Belief Revision

An OCF is a function over possible worlds. The total number of possible worlds is exponential in the number of propositional variables. Therefore, the OCF model for mutual belief revision is not well-suited for providing a computational account of mutual belief revision. In this section, we present another construction of mutual revision operators to investigate computational properties. We will show that the two models are essentially equivalent.

Belief States in Computational Model

We will represent a belief state by an integer-weighted belief base. Formally, an *entrenchment ranking base (ERB)* Ξ consists of a finite set of sentences, B, and a mapping f from B to $\mathbb{N}^+$, which can be represented as $\{\langle\beta_1, e_1\rangle, \cdots, \langle\beta_n, e_n\rangle\}$ such that $\beta_i \in B$ and $f(\beta_i) = e_i$. For any sentence $\beta_i \in B$, we call $f(\beta_i)$ its *evidence degree*. The higher the degree, the firmer is the belief in β_i.

Given an ERB $\Xi = \langle B, f\rangle$, we denote by Ξ^i the subset of B in which all sentences have the same evidence degree i, that is,

$$\Xi^i = \{\beta \in B \mid f(\beta) = i\}$$

Accordingly, $\Xi^{\geq i} = \bigcup\{\Xi^j \mid j \geq i\}$.

The belief set of an ERB $\Xi = \langle B, f\rangle$, denoted by $Bel(\Xi)$, consists of all logical consequence of B, that is, $Bel(\Xi) = Cn(B)$. Much like an OCF, an ERB can also induce a ranking over all sentences:

$$Rank_\Xi(\beta) = \begin{cases} 0, & \text{if } B \not\vdash \beta; \\ \infty, & \text{else if } \vdash \beta; \\ \max\{m \mid \Xi^{\geq m} \vdash \beta\}, & \text{else.} \end{cases} \quad (8)$$

It is easy to show that this defines an AGM epistemic entrenchment ordering.

Note that it is possible that a sentence $\beta \in B$ has a higher belief degree $Rank_\Xi(\beta)$ than its evidence degree $f(\beta)$. So $f(\beta)$ should only be considered as the lower bound of β's epistemic entrenchment ordering. A sentence β is said to be *redundant* in a belief base $\Xi = \langle B, f\rangle$ iff $Rank_\Xi(\beta) > f(\beta)$. Obviously, discarding redundant sentences from a belief base will not affect the rank of any sentence, that is, it will not change the encoded belief state.

An ERB can be related to an OCF. Specifically, we define a function that maps an ERB Ξ to an OCF k_Ξ by setting the rank of a possible world w to the maximal evidence degree of the sentences in the ERB which it does not satisfy:

$$k_\Xi(w) = \begin{cases} 0, & \text{if } w \models B; \\ \max\{f(\beta) \mid \beta \in B \text{ and } w \not\models \beta\}, & \text{otherwise.} \end{cases} \quad (9)$$

The following theorem shows that the OCF k_Ξ induced by an ERB Ξ encodes exactly the same information as Ξ.

Theorem 1. *Suppose* Ξ *is an ERB and* k_Ξ *is the OCF defined by (9). Then for any sentence* β*:*

$$Rank_\Xi(\beta) = Rank_{k_\Xi}(\beta)$$

Revision of ERBs

Similar to the OCF revision operator defined by (4), we propose here an operator which revises an ERB by another ERB.

Suppose $\Xi_1 = \langle B_1, f_1\rangle$ and $\Xi_2 = \langle B_2, f_2\rangle$ are two ERBs. The result of revising Ξ_1 by Ξ_2, denoted by $\Xi_1 \otimes \Xi_2$, is defined as:

$$\begin{aligned} \Xi_1 \otimes \Xi_2 = &\{\langle \beta \vee \alpha, f_1(\beta) + f_2(\alpha)\rangle \mid \beta \in B_1 \text{ and } \alpha \in B_2\} \\ &\cup \{\langle \beta, f_1(\beta) - r\rangle \mid \beta \in B_1\} \cup \Xi_2 \end{aligned} \quad (10)$$

where $r = Rank_{\Xi_1}(\neg \bigwedge B_2)$, assuming $\bigwedge \emptyset = \top$.

The following result establishes the equivalence of our ERB revision operator and the OCF-based revision operator defined by (4).

Theorem 2. *Suppose $\Xi_1 = \langle B_1, f_1 \rangle$, $\Xi_2 = \langle B_2, f_2 \rangle$ are two ERBs and k_{Ξ_1}, k_{Ξ_2} are the OCFs induced by Ξ_1 and Ξ_2, respectively. Then for any possible world w:*

$$(k_{\Xi_1} \otimes k_{\Xi_2})(w) = k_{\Xi}(w)$$

where $\Xi = \Xi_1 \otimes \Xi_2$ and k_{Ξ} is the OCF induced by Ξ.

ERB Model of Mutual Belief Revision

Finally, we define a mutual belief revision operator on ERBs. Technically, this requires the notion of an *i-cut* of an ERB $\Xi = \langle B, f \rangle$, denoted by $\Xi - i$ and defined as follows:

$$\Xi - i = \{\langle \beta, f(\beta) \rangle \mid \beta \in B \text{ and } f(\beta) > i\}$$

Given two ERBs $\Xi_1 = \langle B_1, f_1 \rangle$, $\Xi_2 = \langle B_2, f_2 \rangle$, we define the inconsistency degree, written as $d_{inc}(\Xi_1, \Xi_2)$, of Ξ_1 and Ξ_2 as follows

$$d_{inc}(\Xi_1, \Xi_2) = \begin{cases} 0, & \text{if } B_1 \cup B_2 \nvdash \bot; \\ \max\{i \mid \Xi_1^{\geq i} \cup \Xi_2^{\geq i} \vdash \bot\}, & \text{otherwise.} \end{cases}$$

Definition 3. *The ERB mutual belief revision operator M is defined as follows:*

$$M(\Xi_1, \Xi_2) = \langle \Xi_1 \otimes \gamma_2(\Xi_1, \Xi_2), \Xi_2 \otimes \gamma_1(\Xi_1, \Xi_2) \rangle \quad (11)$$

where $\gamma(\Xi_1, \Xi_2) = \langle \Xi_1 - d_{inc}(\Xi_1, \Xi_2), \Xi_2 - d_{inc}(\Xi_1, \Xi_2) \rangle$.

Let's see some examples to illustrate this operator.

Example 1. *Suppose $\Xi_1 = \{\langle q, 3 \rangle, \langle p, 2 \rangle\}$ and $\Xi_2 = \{\langle q, 3 \rangle \langle \neg p, 2 \rangle\}$. It is easy to see that $d_{inc}(\Xi_1, \Xi_2) = 2$ and $\Xi_1 - 2 = \Xi_2 - 2 = \{\langle q, 3 \rangle\}$. According to (10), we have $M_1(\Xi_1, \Xi_2) = \Xi_1 \otimes (\Xi_2 - 2) = \{\langle q, 6 \rangle, \langle p, 2 \rangle\}$ and $M_2(\Xi_1, \Xi_2) = \Xi_2 \otimes (\Xi_1 - 2) = \{\langle q, 6 \rangle, \langle \neg p, 2 \rangle\}$, after removing redundant sentences.*

In this example, the belief set of each agent remains the same but the evidence degree of the common belief q is increased as a result of the mutual belief revision. Unlike with belief merging, the conflicting opinions of Ξ_1 and Ξ_2 on p are not resolved. It is not difficult to see that the inconsistency remains even if operator M is applied repeatedly.

Example 2. *Suppose $\Xi_1 = \{\langle q, 3 \rangle, \langle p, 1 \rangle\}$ and $\Xi_2 = \{\langle \neg p, 3 \rangle, \langle q, 1 \rangle\}$. In this case we have $d_{inc}(\Xi_1, \Xi_2) = 1$, $\Xi_1 - 1 = \{\langle q, 3 \rangle\}$, and $\Xi_2 - 1 = \{\langle \neg p, 3 \rangle\}$. According to (10), we have $M_1(\Xi_1, \Xi_2) = \{\langle q \vee \neg p, 6 \rangle, \langle \neg p, 3 \rangle, \langle q, 2 \rangle\}$ and $M_2(\Xi_1, \Xi_2) = \{\langle q \vee \neg p, 6 \rangle, \langle q, 4 \rangle, \langle \neg p, 3 \rangle\}$.*

In this example, both agents confirm the common belief $q \vee \neg p$ with high degree. Agent 1 is convinced by agent 2 of $\neg p$ since agent 2 believes in it with high degree and also has a common understanding with agent 1 on q. Agent 2 does not acquire any new belief, but she gets a stronger confirmation in q. As a result of mutual belief revision, the two agents reach a consensus.

$$\begin{tikzcd}
\langle \Xi_1, \Xi_2 \rangle \arrow[r, "(11)"] \arrow[d, "(9)"'] & M(\Xi_1, \Xi_2) \arrow[d, "(9)"] \\
\langle k_{\Xi_1}, k_{\Xi_2} \rangle \arrow[r, "(7)"'] & M(k_{\Xi_1}, k_{\Xi_2})
\end{tikzcd}$$

Figure 2: Equivalence of ERB and OCF-based model

Equivalence Results

To show the equivalence of the two constructions of mutual belief revision presented in this paper, we need the following lemmas.

Lemma 1. *Suppose Ξ is an ERB and k_{Ξ} its induced OCF. Then for any possible world w and natural number r:*

$$(k_{\Xi} - r)(w) = k_{\Xi'}(w)$$

where $\Xi' = \Xi - r$ and $k_{\Xi'}$ is the OCF induced from Ξ'.

Lemma 2. *Suppose Ξ_1, Ξ_2 are two ERBs and k_{Ξ_1}, k_{Ξ_2} are the OCFs induced from Ξ_1 and Ξ_2, respectively. Then*

$$d_{inc}(k_{\Xi_1}, k_{\Xi_2}) = d_{inc}(\Xi_1, \Xi_2)$$

The following equivalence theorem (as illustrated in Figure 2) is a direct consequence of the above lemmas and Theorem 2.

Theorem 3. *Suppose Ξ_1, Ξ_2 are two ERBs and k_{Ξ_1}, k_{Ξ_2} are the OCFs induced from Ξ_1 and Ξ_2, respectively. Then*

$$M(k_{\Xi_1}, k_{\Xi_2}) = \langle k_{\Xi_1'}, k_{\Xi_2'} \rangle$$

where $\langle \Xi_1', \Xi_2' \rangle = M(\Xi_1, \Xi_2)$ and $k_{\Xi_1'}, k_{\Xi_2'}$ are the OCFs induced from Ξ_1' and Ξ_2', respectively.

It follows directly from Theorem 1 and Theorem 3 that the ERB operator shares all of the nice logical properties of the OCF-based operator presented in the previous section.

Computational Complexity

In this section, we present complexity results of two related problems. First of all, we are interested in how hard it is to compute the result of $M(\Xi_1, \Xi_2)$, given two arbitrary ERBs Ξ_1, Ξ_2. It turns out that the problem falls in an interesting complexity class, namely, NP-equivalent, which is the analogue of NP-completeness for function problems. Formally, a function problem is called NP-equivalent iff it is in $\mathrm{F}\Delta_2^{\mathrm{p}}$ (also referred to as NP-easy) and NP-hard.[3]

Theorem 4. *The problem of computing $M(\Xi_1, \Xi_2)$, for arbitrary ERBs Ξ_1 and Ξ_2, is* NP-equivalent.

The second problem is to decide whether an arbitrary sentence β is entailed by both $M_1(\Xi_1, \Xi_2)$ and $M_2(\Xi_1, \Xi_2)$. It turns out that this decision problem inhabits a very low level of the polynomial hierarchy.

[3]Complexity class $\mathrm{F}\Delta_2^{\mathrm{p}}$ (Δ_2^{p}) denotes function (decision) problems which can be solved polynomially by a deterministic Turing machine with an NP-oracle. Readers are referred to (Papadimitriou 1994) for a detailed discussion on these classes.

Theorem 5. *The problem of deciding whether both $M_1(\Xi_1, \Xi_2)$ and $M_2(\Xi_1, \Xi_2)$ entail β, for arbitrary ERBs Ξ_1, Ξ_2 and sentence β, is $\Delta_2^{\mathrm{p}}[\mathrm{O}(\log \mathrm{n})]$-complete.*

Conclusion and Related Work

We have introduced the concept of mutual belief revision. To this end, we have modeled the process of mutual belief revision in two stages: In the first stage, two agents get together trying to reach a common understanding. This stage is quite similar to belief merging (Revesz 1997; Liberatore & Schaerf 1998; Konieczny & Pérez 1998). Once such a common understanding has been reached, two agents revise their belief states in order to incorporate the agreed views into their own belief states. This idea is in the spirit of (Zhang *et al.* 2004).We have introduced two different models for mutual belief revision: an OCF-based model, which clearly shows the intuition and semantics of the operation, and an ERB model suitable for computation.

In this paper, we have focused on situations which involve only two agents. To handle situations with $m \geq 2$ agents, we can generalize γ (as defined by (2)) as follows:

$$\gamma(k_1, \cdots, k_m) = \langle k_1 - r_n, \cdots, k_m - r_n \rangle$$

where $r_n = \min\{r \mid k_1^-(r) \cap \cdots \cap k_m^-(r) \neq \emptyset\}$. Accordingly, M (as defined by (7)) can be extended to

$$M(k_1, \cdots, k_m) = \langle k_1 \otimes T_1, \cdots, k_m \otimes T_m \rangle$$

where $T_i = \bigoplus_{1 \leq j \leq m, j \neq i} \{\gamma_j(k_1, \cdots, k_m)\}$.[4]

In (Zhang *et al.* 2004), a formalism is presented in which negotiation is viewed as a process of mutual belief revision. In this work, a set of AGM-style postulates are proposed for mutual belief revision operators. However, no semantic model or computational model is given. Moreover, negotiation and mutual belief revision should not be considered interchangeable despite some similarities in their operation. The focus of negotiation is to reach a mutually beneficial agreement; once such an agreement is formed, the divergence of beliefs among the agents is no longer of importance. In contrast, the divergence of beliefs is always the focus of mutual belief revision, before and after the process. In this sense, negotiation is closer to belief merging than mutual belief revision.

An interesting perspective for future work is of course to formulate AGM-style postulates for mutual belief revision. As a staring point, we might directly translate some properties of Proposition 3 into AGM-style postulates and provide a representation theorem for these postulates. As an example, property (M3) corresponds directly to so-called *consistent expansion* of (Zhang *et al.* 2004), which can be reformulated as follows:

- If K_i and $M_i(K_1, K_2)$ are consistent, then $K_i \subseteq M_i(K_1, K_2)$.

The main difference between mutual belief revision and the negotiation operation of (Zhang *et al.* 2004) is that the former does not satisfy the following so-called *no recantation* postulate.

- $K_2 \cap M_1(K_1, K_2) \subseteq M_2(K_1, K_2)$
  $K_1 \cap M_2(K_1, K_2) \subseteq M_1(K_1, K_2)$

At first glance, this postulate seems quite natural: an agent should not give up its own beliefs that will be accepted by the other agent. However, as shown in (Meyer *et al.* 2004b; 2004a), no recantation also enforces the participating agents to have identical beliefs after negotiation. Therefore, we argue that no recantation is too strong a general postulate for mutual belief revision.

It is also worth mentioning that both projections (i.e., M_1 and M_2) of mutual belief revision satisfy most properties of so-called selective revision (Fermé & Hansson 1999). In particular, Property (M4) guarantees the satisfiability of Postulate *proxy success*, which can be reformulated as follows:

- There is a subset Y of X, such that $Y \subseteq K * X$ and $K * X = K * Y$.

[4]We denote by $k_1 \oplus k_2$ the *expansion* of k_1 by k_2, such that $(k_1 \oplus k_2)(w) = k_1(w) + k_2(w)$. Note that $\oplus$ is commutative, and $k_1 \oplus k_2 = k_1 \otimes k_2$ when k_1 and k_2 are mutually consistent.

References

Darwiche, A., and Pearl, J. 1997. On the logic of iterated belief revision. *Artificial Intelligence* 89:1–29.

Fermé, E. L., and Hansson, S. O. 1999. Selective revision. *Studia Logica* 63(3):331–342.

Gärdenfors, P., and Makinson, D. 1988. Revisions of knowledge systems using epistemic entrenchment. In *Proceedings of TARK: Theoretical Aspects of Reasoning about Knowledge*, 83–95.

Grove, A. 1988. Two modellings for theory change. *Journal of Philosophical Logic* 17:157–180.

Hansson, S. O. 1999. A survey of non-prioritized belief revision. *Erkenntnis* 50:413–427.

Jin, Y., and Thielscher, M. 2007. Iterated belief revision, revised. *Artificial Intelligence* 171(1):1–18.

Konieczny, S., and Pérez, R. P. 1998. On the logic of merging. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 488–498.

Liberatore, P., and Schaerf, M. 1998. Arbitration (or how to merge knowledge bases). *IEEE Transactions on Knowledge and Data Engineering* 10(1):76–90.

Meyer, T.; Foo, N.; Kwok, R.; and Zhang, D. 2004a. Logical foundations of negotiation: Outcome, concession, and adaptation. In *Proceedings of AAAI*, 293–298.

Meyer, T.; Foo, N.; Kwok, R.; and Zhang, D. 2004b. Logical foundations of negotiation: Strategies and preferences. In *Proceedings of KR*, 311–318.

Papadimitriou, C. 1994. *Computational Complexity*. Addison-Wesley.

Revesz, P. Z. 1997. On the Semantics of Arbitration. *Journal of Algebra and Computation* 7(2):133–160.

Spohn, W. 1988. Ordinal conditional functions: A dynamic theory of epistemic state. In Harper, W. L., and Skyrms, B., eds., *Causation in Decision: Belief, Change and Statistics: Proceedings of the Irvine Conference on Probability and Causation*, volume II, 105–134. Dordrecht: Kluwer Academic Publisher.

Spohn, W. 1991. A reason for explanation: Explanations provide stable reasons. In Spohn, W., van Fraassen, B.C., and Skyrms, B., eds., *Existence and Explanation*. Kluwer Academic Publisher. 165–196.

Zhang, D.; Foo, N.; Meyer, T.; and Kwok, R. 2004. Negotiation as mutual belief revision. In *Proceedings of AAAI*, 317–322.