Multi-Agent Belief Revision

Noa E. Kfir-Dahav and Moshe Tennenholtz
Faculty of Industrial Engineering and Management
Technion – Israel Institute of Technology
Haifa 32000, Israel

Abstract

We introduce a basic framework for multi-agent belief revision in heterogeneous societies where agents are required to be consistent in their beliefs on shared variables. We identify several properties one may require a general multi-agent belief revision operator to satisfy, and show several basic implications of these requirements. Our work reveals the connection between multi-agent belief revision and the theory of social choice, and attempts to provide some initial understanding of the multi-agent belief revision process.

1 Introduction

Consider an agent, a set of beliefs it adopts, and a new observation it makes. Taking the new observation as a fact that should be incorporated to its beliefs, the agent may need to revise its beliefs. Belief revision is a basic research topic in AI [Gar88, KM91, FH94, dS92, Gro88, Doy91], and research on the theory of belief revision seems to converge to a well-established theory [Gar92]. Most of the theory developed in this area has concentrated on systems consisting of a single agent. Our aim in this paper is to use the current understanding of the single agent belief revision process in order to discuss the more general process of multi-agent belief revision.

Consider a set of agents operating in a shared environment. The agents may be robots, databases, or other artificial entities. Each of them may have its own perspective on the world, but usually there will be several elements that are shared among the agents. For example, in the context of distributed databases, different agents may have different data items they refer to, but there are also common data items that several agents may refer to. Assuming that the perspective of an agent is captured by a set of beliefs, and that each agent may encounter new observations, the agents may need to revise their beliefs.

Most previous work on multi-agent belief revision has emphasized the revision of an agent's model of other agents' models. For example, in [vdM94] Meyden considers the case where new observations are observed by all agents simultaneously, and as a result of such new observation agents revise their beliefs about the beliefs of other agents. Other researchers (e.g., Gaspar [Gas91] and Galliers [Gal92]) consider situations where observations made by an individual agent may change its beliefs and its beliefs about the beliefs of other agents, as well as effect the content of messages the observer may wish to send to these agents. In this paper we wish to take another perspective on multi-agent belief revision. This paper initiates research on multi-agent belief revision in heterogeneous systems, where each agent may have its own perspective of the world but the agents need to coordinate (i.e., agree on) their beliefs on shared elements. Unfortunately, excluding some work on algorithmic aspects of multi-agent belief revision [HB91], no work has addressed this general heterogeneous multi-agent belief revision process. However, the latter is crucial for many applications, such as in the context of distributed databases.

In this work we introduce a basic framework for multi-agent belief revision, discuss several requirements one may require a satisfactory multi-agent belief revision to satisfy, and present some basic implications of these requirements. In particular, our work reveals a connection between multi-agent belief revision and the theory of social choice, a connection we believe to be of significant importance.

The structure of this paper is as follows. In Section 2 we briefly discuss part of the theory regarding single agent belief revision which we will make use of in later sections. In Section 3 we introduce the multi-agent belief revision setting. Section 4 lists our set of requirements from the multi-agent revision process. We consider consistency and rationality requirements, social rationality requirements, and additional semantic requirements. While these requirements are not ultimate, we believe they capture the type of the properties a satisfactory multi-agent revision operator should have. In Section 5 we present the implications of our requirements. In particular, we present several basic results regarding the existence and uniqueness of a satisfactory multi-agent belief revision operator. In Section 6 we further discuss our setting and results, as well as some related work.

2 Preliminaries

The discussion in the following sections will make use of some basic notions and results of the theory of single agent belief revision. This section presents a brief introduction to these notions and results.

Definition 2.1: Let $\Phi = \{\varphi_1, \ldots, \varphi_n\}$ be a set of primitive propositions. Let $\mathcal{L}$ be the closure of Φ under $\neg, \vee$. A *belief* is a sentence $KB \in \mathcal{L}$. Alternatively, a belief $KB \in \mathcal{L}$ is associated with its set of truth assignments.[1] Given a current belief KB and a new observation O, a *belief revision* is a transformation from $KB \in \mathcal{L}$ to $KB' \in \mathcal{L}$ where all truth assignments of KB' are truth assignments of O.

A belief revision is therefore a process in which an agent changes its current beliefs in a way that the set of truth assignments associated with the new beliefs will also be truth assignments of the new acquired observation. Naturally, not all belief revision operators are satisfactory. A satisfactory (single agent) belief revision operator is assumed to obey the AGM postulates [AGM85, Gar88]. These postulates aim to capture stability properties, eliminating unnecessary perturbations to the database. For example, one of these postulates states that if a new acquired observation is consistent with the agent's beliefs, then the agent's new belief is obtained by adding the new observation to the current beliefs. In the case of a propositional knowledge base, it has been shown [KM91] that the AGM postulates coincide with the following conditions (the Katsuno and Mendelzon conditions), where KB is the current belief, O is the current observation, and $KB \circ O$ is the result of revising KB by O:

[1] For a discussion of various approaches to modeling belief and belief change the reader may consult [Gro88].

1. $KB \circ O$ implies O
2. If $KB \wedge O$ is satisfiable then $KB \circ O \equiv KB \wedge O$.
3. If O is satisfiable then $KB \circ O$ is also satisfiable.
4. If $KB_1 \equiv KB_2$ and $O_1 \equiv O_2$ then $KB_1 \circ O_1 \equiv KB_2 \circ O_2$
5. $(KB \circ O) \wedge O'$ implies $KB \circ (O \wedge O')$
6. If $(KB \circ O) \wedge O'$ is satisfiable then $KB \circ (O \wedge O')$ implies $(KB \circ O) \wedge O'$

In the sequel we will refer to the above conditions as the KM conditions. The above conditions capture basic properties required from a satisfactory belief revision operator. In addition, Katsuno and Mendelzon supply the following representation theorem:

Theorem 2.2: *A belief revision operator $\circ$ satisfies the Katsuno and Mendelzon conditions if and only if there is a persistent assignment* [2] *which maps each belief KB to a total pre-order $\leq_{KB}$ over the possible assignments to Φ so that the truth assignments to $KB \circ O$ are the minimal truth assignments of O with respect to $\leq_{KB}$.*

In this paper we will not provide the details of the above basic result. However, the message of this result, as well as of other results which followed and extended it (e.g., [FH94]) is both simple and powerful. Roughly speaking, it can be summarized as follows. A satisfactory belief revision operator can be viewed as an operator which generates a ranking (i.e., a total pre-order) over the set of possible assignments to Φ as a function of the current belief, where the set of minimal elements in this ranking coincides with the current belief. Given a new observation O the operator will choose the minimal assignments in this ranking that satisfy O to be the new (revised) belief. As a result, we can view a single agent belief revision operator as a mapping $(R^{old}, O) \rightarrow R^{new}$ where R^{old} and R^{new} are rankings over the set of possible assignments to Φ.

[2]The definition of persistent assignment and other details regarding this result are omitted from this paper. For full details see [KM91].

3 Multi-Agent Belief Revision

In this section we define a basic setting where multi-agent belief revision can be investigated. We treat multi-agent belief revision as a two-phase process and discuss our strategy for defining satisfactory multi-agent belief revision operators. In this paper we will assume that there are only two agents in our system.[3]

Definition 3.1: Let $A = \{1, 2\}$ be a set of *agents* functioning in a partially shared *environment*. The environment is described by a set of primitive propositions $\Phi = \{\varphi_1, \ldots, \varphi_n\}$. For each agent i there is a set $\Phi_i \subset \Phi$ of primitive propositions that describe agent i's view of the environment. A proposition $\varphi \in \Phi$ is a *private proposition* of agent i if $\varphi \in \Phi_i \setminus \Phi_j$ for $j \neq i$. A proposition $\varphi \in \Phi$ is a *shared proposition* if $\varphi \in \Phi_i \cap \Phi_j$ for $j \neq i$. The *private domain of agent i* consists of all the private propositions of this agent, while the *shared domain* of the agents consists of all of their shared propositions. An *observation* by agent i is a set of possible truth assignments to a subset of its private propositions or to a subset of the shared propositions.

The above definition captures a general setting where each agent has private propositions as well as other propositions which are shared with other agents. The shared domain defines also the communication language for the agents. Agents will be able to communicate only about elements of the shared domain. We make the assumption that an observation refers either to private or to shared propositions, and that the shared domain includes at least three primitive propositions. We will also make the assumption that only one agent can make an observation at a given point.[4]

Given a new observation the agents would need to incorporate it into their beliefs. Our main question is: what would be the structure of a belief revision process in a multi-agent setting?

[3]Our results can be extended to some n-party settings, where $n > 2$. We will return to this point in Section 6. This point is discussed in detail in [KD96].

[4]Similar settings can be defined for the case where agents may have simultaneous observations, or observations about a mixture of private and shared propositions. Although the belief revision operators which the agents use may be defined for these cases as well, our multi-agent setting assumes the above-mentioned restrictions.

Given the above setting we view the multi-agent belief revision as a two-phase process. In the first phase, if the new observation contains information regarding Φ_i then agent i performs an individual belief revision. We assume that an observation about the shared domain is communicated by the observer to the other agent; hence, in the latter case both agents would make an individual belief revision. In the second phase the agents may wish to coordinate their beliefs. The first phase is in fact a single-agent belief revision process as investigated in previous work; the second phase is still an unexplored step which should be carefully investigated.

Given the above discussion we can define:

Definition 3.2: A *multi-agent belief revision* is a mapping

$$MULTI-REV : (KB_1^{old}, KB_2^{old}, O, a) \rightarrow (KB_1^{new}, KB_2^{new}),$$

where KB_i^{old} is the old belief of agent i, KB_i^{new} is the new belief of agent i, O is an observation, and a is the identity of the agent who made the observation. The *individual belief revision phase* is the first phase of the multi-agent belief revision process, and it is a mapping $Step_1 : (KB_1^{old}, KB_2^{old}, O, a) \rightarrow (KB_1^I, KB_2^I)$, where $KB_i^I = KB_i^{old} \circ O$ if the value of a is i or O refers to the shared domain, and $KB_i^I = KB_i^{old}$ otherwise.

Assuming the Katsuno and Mendelzon conditions hold for the individual belief revision phase, we get that this phase can be treated as a mapping $Step_1 : (R_1^{old}, R_2^{old}, O, a) \rightarrow (R_1^I, R_2^I)$, where R_i^{old} and R_i^I are the rankings associated with agent i before and after the individual belief revision phase respectively.

Given the above discussion we are still left with a degree of freedom relative to the second phase of the multi-agent belief revision process. Using the terminology of rankings, we can treat the second phase of the multi-agent belief revision process as a mapping $Step_2 : (R_1^I, R_2^I, O, a) \rightarrow (R_1^{new}, R_2^{new})$, where R_i^I is as defined above and R_i^{new} is the ranking based on which individual belief revision will be performed on the *next* belief revision iteration.

Notice that the above discussion does not assume that an agent needs to manipulate explicit rankings. All we assume is that agents perform a two-phase belief revision process, where the first phase is a single agent belief revision process which obeys classical belief revision postulates. In order to have a full characterization of multi-agent belief revision, we would need to add a list of requirements from $Step_2$ of the multi-agent belief revision process. This will be the topic of the following section.

4 The Multi-Agent Belief Revision Requirements

In this section we discuss the second phase of the multi-agent belief revision process that we presented in the previous section. Motivated by the work on single agent belief revision, our aim is to suggest an initial set of requirements from the multi-agent belief revision process and to investigate what do these requirements entail. For technical reasons, and for ease of exposition, we will present the set of requirements in four stages.

In the sequel we will refer only to the requirements from the second (joint) phase of the multi-agent belief revision process. Recall that this process can be described by the mapping $Step_2$ defined in the previous section.

4.1 Consistency

In this work we are concerned with heterogeneous systems where agents wish to coordinate their activities. In particular, they would need to be consistent about their beliefs on the shared domain. Without such an agreement, the system will lack sufficient coordination.

Definition 4.1: Let KB_i be the belief, interpreted as a set of truth assignments, of agent i. Let us denote by $KB_i(\Phi_1 \cap \Phi_2)$ the projection of KB_i on the shared domain. This projection is defined as the union of the projections of the truth assignments of KB_i on $\Phi_1 \cap \Phi_2$.

Requirement 1 (Consistency): Let KB_i^{new} be the belief of agent i by the end of the second phase of a multi-agent revision process. Then, $KB_1^{new}(\Phi_1 \cap \Phi_2) = KB_2^{new}(\Phi_1 \cap \Phi_2)$.

The above requirement states that the multi-agent belief revision process should ensure that agents will be consistent in their beliefs on the shared domain. For ease of exposition we assume that the beliefs of the agents are initially consistent.

Given the consistency requirement we can view $Step_2$ as a two-step mapping:

$$Step_{2.1} : (R_1^I, R_2^I, O, a) \rightarrow \psi$$

$$Step_{2.2} : (R_1^I, R_2^I, \psi, O, a) \rightarrow (R_1^{new}, R_2^{new})$$

In the above mappings ψ denotes a set of assignments to the shared domain which the agents agree upon. We require that $KB_1^{new}(\Phi_1 \cap \Phi_2) = KB_2^{new}(\Phi_1 \cap \Phi_2) = \psi$. Notice that $Step_{2.2}$ is, yet again, a single-agent belief revision process where agents need to revise their beliefs to be consistent with ψ. However, $Step_{2.1}$ is a subtle step which should be carefully discussed.

4.2 Rationality

By now we have a general setting of multi-agent belief revision where agents are required to be consistent in their beliefs on the shared domain. However, consistency does not suffice to obtain coordinated behavior. In particular, it is only natural to require that the agents will behave as a single agent as far as belief revision based on observation about the shared domain is concerned.[5] Using the machinery of Section 2 it is easy to see that this common-sensical requirement can be translated into the following pair of requirements, which we call *rationality requirements.* Notice that these requirements enforce a slight change in the definition of $Step_{2.1}$.

[5]The requirements from single-agent belief revision implies that a single ranking will be associated with each belief of an agent; however, this is an unnatural requirement when we consider only beliefs over the shared domain. Hence, in our setting the agents will need to behave as an agent with a particular belief revision operator about the shared domain (which they need to implicitly agree upon at each point as we describe), but this belief revision operator may change in the following iterations. This point is further discussed in [KD96] and in the full paper. We will return to this point at Section 6

Requirement 2: The output of $Step_{2.1}$ is a joint ranking over the possible assignments to $\Phi_1 \cap \Phi_2$.

Requirement 3: If the observation made is about the shared domain, then the belief on the shared domain will be revised based on the previously agreed upon ranking over the possible assignments to $\Phi_1 \cap \Phi_2$.

Notice that the above requirements are again a straightforward interpretation of the classical single-agent belief revision theory, when applied to the shared domain. Given the above requirements, we can redefine the multi-agent belief revision process to consist of the following steps:

$$Step_1 : (R_1^{old}, R_2^{old}, R_s^{old}, O, a) \rightarrow (R_1^I, R_2^I)$$

$$Step_{2.1} : (R_1^I, R_2^I, R_s^{old}, O, a) \rightarrow (R_s^{new})$$

$$Step_{2.2} : (R_1^I, R_2^I, R_s^{new}, O, a) \rightarrow (R_1^{new}, R_2^{new}, R_s^{new})$$

In the above mappings R_s^{old} and R_s^{new} denote the previously agreed upon ranking over the shared domain, and the new agreed upon ranking over the shared domain, respectively. Initially, these rankings are undefined. Notice that the set of minimal elements in R_s^{new} and the sets determined by projecting the minimal elements in R_1^{new} and R_2^{new} on the shared domain, should be identical. In a case where the observation O is about the shared domain, $Step_{2.1}$ would require that the minimal elements in R_s^{new} would coincide with the minimal elements is R_s^{old} that satisfy O. We would like to emphasize again that what we obtained here is a characterization of multi-agent belief revision by means of mappings between rankings. These however may be implicit rankings; the key point is that we can view the multi-agent belief revision process **as if** it consists of a mapping between rankings as defined above.

Notice that, given the above structure, the multi-agent belief revision process is an **iterative** process which is initiated each time a new observation is made. *Each iteration* of this process can be *viewed* as if it consists of the previously mentioned stages. For ease of presentation, we will assume that the first observation made (when the system is initiated) is about the private domain of one of the agents.

4.3 Social Rationality

Our previous discussion left us with a degree of freedom relative to $Step_{2.1}$. This step should determine the new joint ranking of the agents over the shared domain. Naturally, this new ranking should be a function of the individual rankings of the agents over the shared domain in the beginning of $Step_{2.1}$.

The individual ranking of agent i over the shared domain, $R_i^{I,S}$, can be extracted from its individual ranking R_i^I as follows. Let W denote the set of possible assignments to Φ_i, and let W_s denote the set of possible assignments to $\Phi_1 \cap \Phi_2$. For every $w_s \in W_s$, let $G(w_s) \in W$ be a minimal element according to R_i^I which projects w_s on the shared domain. We can now define the ranking $R_i^{I,S}$ by defining $w_s \leq w'_s$ to hold if and only if $G(w_s) \leq G(w'_s)$ in R_i^I.

The mapping defined in $Step_{2.1}$ should take into account the individual rankings of the agents, since these rankings capture the private beliefs of the agents. Therefore, we should require *social rationality* about the way R_s^{new} is built given $R_1^{I,S}$ and $R_2^{I,S}$. Requirements for social rationality can be found in the social choice literature [LR57, Arr63]. These requirements were already used in the AI literature in a somewhat similar context [DW89].

We will adopt the following basic social rationality requirements.

Requirement 4 (Generality): The mapping defined in $Step_{2.1}$ should be defined for all $R_1^{I,S}$ and $R_2^{I,S}$.

Requirement 5 (Weak Pareto Optimality): For every $x, y \in W_s$, if $x < y$ both in $R_1^{I,S}$ and in $R_2^{I,S}$ then $x < y$ should also hold in R_s^{new}.

Requirement 6 (Independence of Irrelevant Alternatives): For every $x, y \in W_s$, the relation between x and y in R_s^{new} depends only on their relation in $R_1^{I,S}$ and $R_2^{I,S}$.

4.4 Semantic Requirements

The last set of requirements relates to the fact the process defined above should be a fair, adequate, and cautious belief revision process.

Requirement 7: The result of the multi-agent belief revision process should be independent of the names of the agents (i.e., 1 and 2).

Requirement 8: If an observation about the private domain of agent i has changed its ranking over the shared domain then R_s^{new} should be different from R_s^{old}.

Requirement 8 may seem a bit technical, but it captures the desire that changes caused by private observations cannot be simply neglected. Notice that the need to consider shared observations has already been addressed by other requirements.

Requirement 9: Let R_s^{new1} and R_s^{new2} be two candidates for R_s^{new} in a particular iteration of a multi-agent belief revision process which satisfies requirements 1–8. Assume that R_s^{new1} and R_s^{new2} coincide until the k'th rank, but the k-th rank of R_s^{new1} strictly contains the k-th rank of R_s^{new2}. Then, the process should prefer to choose R_s^{new1} rather than R_s^{new2}.

Requirement 9 captures the need to be as cautious as possible in ignoring some of the possible assignments, which is a typical assumption in non-monotonic reasoning and belief ascription contexts [BT94]. Although this requirement is natural in many domains, one can consider situations in which this requirement is not essential for defining satisfactory multi-agent belief revision operators. Results similar to the ones presented in the following section can be obtained if we drop requirement 9 or replace it by similar requirements.

5 Implications of the Multi-Agent Belief Revision Requirements

The previous section defined a set of requirements from the multi-agent belief revision process. Although we believe that these requirements are not ultimate ones, we think they capture the type of requirements a satisfactory multi-agent revision operator should satisfy. Hence, it may be of considerable importance to study the implications of these requirements on the identity of a satisfactory multi-agent belief revision operator.

First, we can show that:

Theorem 5.1: *Any mapping of $R_1^{I,S}$ and $R_2^{I,S}$ to R_s^{new} which satisfies requirements 2,4-6, and 9, would have the property that $R_s^{new} = R_1^{I,S}$ or $R_s^{new} = R_2^{I,S}$.*

The above theorem implies that given two individual rankings over the shared domain in $Step_{2.1}$, a mapping satisfying our requirements would output one of these rankings as R_s^{new}. This result is mainly an implication of Arrow's Impossibility Theorem [Arr63]. Notice that the process we consider consists of a sequence of iterations, where in each iteration a multi-agent belief revision operator is applied. The above theorem implies that given a particular iteration of the multi-agent belief revision process, if we denote the individual rankings of the agents over the shared domain in $Step_{2.1}$ by $Rank_1$ and $Rank_2$ respectively, then R_s^{new} will be either $Rank_1$ or $Rank_2$. Given this we can show that:

Theorem 5.2: *Assuming the agents have fixed syntactic names, there is no multi-agent belief revision process satisfying requirements 1–9. Moreover, there is no multi-agent belief revision process satisfying requirements 1–6 and 8–9.*

The above theorem refers to the case where we use fixed syntactic names for the agents. However, one may think of semantic names, where the names of the agents are not fixed, but are determined by the system's history. For example, one may wish to say that as far as joint decisions are concerned, agent 1 is taken to be the agent who observed the latest observation. The syntactic names of the agents will then play no role in joint decisions. Such semantic names can be captured by a *naming* function which maps $R_1^{I,S}$ and $R_2^{I,S}$ to $Rank_1$ and $Rank_2$ based on the previous history of the agents.[6] The requirements about the connection between $R_1^{I,S}$ and $R_2^{I,S}$ to the new joint ranking, will be replaced by similar requirements about the connection between $Rank_1$ and $Rank_2$ to the new joint ranking. The case where $Rank_i = R_i^{I,S}$ for $i = 1, 2$, independently of the previous history, gives us the case of fixed syntactic names.

[6]Naturally, we may wish to consider only the part of the history which both agents can refer to.

In the sequel, the history of agents will include the sequence of previous beliefs on the shared domain and observations made. The naming function will be applied at each iteration to the agents' joint history. Formally, this history is a sequence $(B_0, Ob_1, B_1, Ob_2, B_2, \ldots, B_{n-1}, Ob_n)$, where B_i is the agents' belief on the shared domain after the i's observation, and Ob_i is the identity of the agent who made the i-th observation, where Ob_n refers to the identity of the last agent to initiate a belief revision iteration; the history also stores the type (i.e., private or shared) of each Ob_i; for each Ob_i $(0 < i < n)$ which refers to an observation about the shared domain, the history may also store the value of the corresponding observation. Naturally, the naming function will not refer to the content of private observations but only to the identity of their observer.

Recall that in our study we refer to a sequence of iterations where each iteration consists of a single multi-agent belief revision process. The naming function is the tool by which fairness requirements and other semantic considerations can be better incorporated into the system. Given that there is an agent who would dictate the new ranking in a given iteration, the naming function can determine the identity of the dictator.

We can show that:

Theorem 5.3: *Let $Rank_1 = R_i^{I,S}$ where i is the agent who made the last observation. Then, there is no multi-agent belief revision process satisfying requirements 1-9.*

Theorem 5.4: *Let $Rank_1 = R_i^{I,S}$ where i is the agent who made the last* private *observation (i.e., observation about a private domain). Then, there exists a multi-agent belief revision process satisfying requirements 1-9. This multi-agent belief revision satisfies that $R_s^{new} = Rank_1$ at each iteration.*

The above theorems show that if the naming function always selects as agent 1 (or always selects as agent 2) the agent who was the last to make an observation, then we will not get a satisfactory multi-agent belief revision; however, if the naming of agents will change only after having private observations then we will get the existence of a satisfactory multi-agent belief revision process. Moreover, we can show that the above naming function is the *only* one which will yield a satisfactory multi-agent belief revision process:

Theorem 5.5: *The naming function defined in Theorem 5.4, is the only naming function under which requirements 1–9 are satisfied.*

The above theorems suggest to us a particular satisfactory multi-agent belief revision operator. According to this operator the revision process will be carried out by the agent which was the last to make a private observation, using the original belief revision operator of this agent. The outcome of the revision made by this agent will be announced to the other agent, which will then revise its beliefs appropriately. Notice that this multi-agent belief revision operator is easily implementable. Each agent will need to keep track of which agent has made the last private observation, but the revision itself will be carried out using the single agent belief revision operators the agents are occupied with.

6 Discussion

In the previous sections we have introduced a basic multi-agent belief revision setting, and presented some basic requirements from a multi-agent belief revision process. We have shown several implications of these requirements; these include results pertaining the existence and uniqueness of a satisfactory multi-agent belief revision operator. In this paper we made some assumptions which can be relaxed. For example, one may consider different variants of the rationality requirement, and of the pareto-optimality assumption. Nevertheless, simple variants of these requirements can be shown to lead to similar results. However, it is not our claim that our requirements are ultimate. Our belief however is that our requirements and results supply an initial rigorous setting that enables the initiation of research on multi-agent belief revision in heterogeneous societies.

The reader should notice that we have not required that the agents will behave as if they have a fixed belief revision operator over the shared domain. At each iteration, a revision of the joint beliefs over the shared domain will satisfy the KM conditions; however, if we consider the process as a whole (i.e., require the KM conditions regarding revisions that may be made in different iterations) then condition 4 of the KM conditions may not be satisfied. This is implied by the fact that we view step 2.1 of each iteration, as a decision about the joint belief revision operator (over the shared domain) for the following iteration. Naturally, this decision may depend on the whole information the agents have, and not only on their decision on joint beliefs over the shared domain.

In this paper we have restricted ourselves to the case of two agents. One may consider a larger group of agents. Our results can be extended to the case where a fixed number of agents share a single shared domain (i.e., each primitive proposition is either associated with only one of the agents or is associated with all of the agents). In this case the multi-agent belief revision process and the corresponding requirements are similar to the ones which we have used in the case of two agents, and lead to similar results. If different agents may have different shared domains then the problem is much more complicated and further study seems to be required.

We mentioned some related work in the body of this paper. We should emphasize that our work has been inspired by and is complementary to several lines of research. First, it is related to work on applying choice theory to the AI context [DW89, Doy91]. Second, it is related to work on the general theory of multi-agent belief revision for uncoordinated[7] agents [vdM94, Gal92, Gas91] , and to the study of algorithmic aspects of multi-agent belief revision [HB91]. Less directly, this work is related to work on multi-agent non-monotonic reasoning [Mor90] as well as to work on speech acts [GS90, AK88]. Last but not least our work heavily relies on the strong foundations supplied by theories of single-agent belief revision [Gar88, KM91, FH94, dS92, GM88, Bou92, Gro88], which enable us to treat the single-agent revision process as a building block for the multi-agent case. Our contribution is the introduction of basic principles and results for the general theory of multi-agent belief revision where agents are required to be both rational and consistent in their revisions.

Appendix: Sketch of Proofs

Proof of Theorem 5.1:

[7]See our discussion in the introduction of this paper.

In order to satisfy requirement 2, R_s^{new} should be a joint ranking over the possible assignments to $\Phi_1 \cap \Phi_2$. Requirements 4-6 coincide with the conditions of Arrow's Impossibility Theorem [Arr63]. In Arrow's setting, one wishes to generate a joint ranking from a set of individual rankings in a way that will be general, (weak) pareto-optimal, and independent of irrelevant alternatives. Arrow's Impossibility Theorem implies that any mapping from a set of rankings to a joint ranking, which satisfies these requirements, will be a dictatorial mapping. In a dictatorial mapping there is an agent such that for every pair (x, y) of assignments to $\Phi_1 \cap \Phi_2$, if this agent strictly prefers x over y then x will be strictly preferred over y in the joint ranking. This implies that in any mapping from $(R_1^{I,S}, R_2^{I,S})$ to a joint ranking, which satisfies requirements 2,4-6, one of the agents' individual rankings (e.g., $R_1^{I,S}$ or $R_2^{I,S}$) will be the basis of the joint ranking; I.e., every strict preference in this ranking will be imposed on the joint ranking. At the most, the mapping will allow the other agent to influence the order of assignments among which the dictatorial agent is indifferent. However, Requirement 9 implies that in generating the joint ranking we will prefer to adopt each rank of the dictator as is. ∎

Proof of Theorem 5.2:

The fact that requirements 1–9 are not simultaneously satisfied is an immediate corollary of Theorem 5.1. The proof of Theorem 5.1 also implies w.l.o.g that agent 1 will dictate the joint ranking at each iteration. As a result, a private observation by agent 2 which changes its beliefs on the shared domain will not change the joint ranking. This contradicts requirement 8. ∎

Proof of Theorem 5.3:

There are two possible mappings of $(Rank_1, Rank_2)$ to R_s^{new}. If $R_s^{new} = Rank_1$ then the dictatorial agent at each iteration is the agent who made the last observation. If $R_s^{new} = Rank_2$ then the dictatorial agent at each iteration is the agent who did not make the last observation.

Consider the case where $R_s^{new} = Rank_1$. Assume that the first observation is a private observation made by agent 1, and that the second observation is a shared observation made by agent 2. Given that $R_s^{new} = Rank_1$, agent 1 will determine the joint ranking of the agents at the end of the first iteration, and agent 2 will determine the outcome at the end of the second iteration. Notice that Requirement 3 requires that given an observation about the shared domain the beliefs over the shared domain will be revised based on R_s^{old}. However, since agent 1 is the dictator of the first iteration, and it dictates only the beliefs in the beginning of the second iteration (and not the agents' individual rankings in this stage) we get that Requirement 3 can not in general be satisfied.

Consider the case where $R_s^{new} = Rank_2$. In this case it is easy to see that a private observation will never affect the joint ranking and requirement 8 is not satisfied. ∎

Proof of Theorem 5.4:

In $step_{2.2}$ the agent which is not the dictator should revise its beliefs to coincide with the dictator's beliefs over the shared domain. This implies that Requirement 1 is satisfied.

The outcome of $step_{2.1}$ is $R_i^{I,S}$ where i is the agent who made the last private observation. $R_i^{I,S}$ is a ranking over the possible assignments to $\Phi_1 \cap \Phi_2$; therefore, the outcome of $step_{2.1}$ is a ranking over the possible assignments to $\Phi_1 \cap \Phi_2$, and Requirement 2 is satisfied.

In an iteration initiated by an observation over the shared domain, the dictator is the dictator of the previous iteration. Without loss of generality, if agent 1 was the dictator of the previous iteration then in the previous iteration $R_s^{new} = R_1^{I,S}$ and agent 1 needs not perform additional revision in $step_{2.2}$. Given that no change of dictator can occur between that (previous) iteration and the current one, the new revision will be made based on the individual ranking over the shared domain of agent 1, which coincides with the agreed upon ranking of the previous stage. Hence, Requirement 3 is satisfied.

The mapping in $step_{2.1}$ is a fully dictatorial mapping, which uses only a semantic consideration when deciding on the identity of the dictator. Hence, it is easy to check that Requirements 4-7 and 9 are satisfied.

If agent i makes a private observation then it determines the joint ranking. Hence, the effects of the new observation on the agents' view of the shared domain are fully reflected in R_s^{new}, and Requirement 8 is satisfied. ∎

Proof of Theorem 5.5:

The proof follows from the following pair of lemmas.

Lemma 1: Any multi-agent belief revision process that satisfies requirements 2-6 and 9 will be a process in which the dictator under a shared observation has been also the dictator of the previous iteration.

Proof of Lemma 1: As we have shown in the proof of Theorem 5.1, if the process satisfies requirement 2,4-6 and 9 then $R_s^{new} = Rank_1$ or $R_s^{new} = Rank_2$. Requirement 3 demands that if we are revising according to a shared observation, the minimal assignments according to R_s^{new} will be identical to the minimal assignments according to R_s^{old} which satisfy the new observation. The only way to ensure this (without relating to the agents' current beliefs and observation or to the complete previous rankings) is to have as the dictator the agent whose individual ranking over the shared domain by the end of the previous iteration coincides with R_s^{old}. This agent needs to be the dictator of the previous iteration.

Lemma 2: Any multi-agent belief revision process that satisfies requirements 2,4-6,8 and 9 will be a process in which the dictator under a private observation is the agent who made the current observation.

Proof of Lemma 2: If the process satisfies requirement 2,4-6 and 9 then $R_s^{new} = Rank_1$ or $R_s^{new} = Rank_2$. Requirement 8 demands that an agent who made a private observation will be able to influence the joint ranking and will not be ignored. Any mapping that will allow in some point for a non-observer to dictate the joint ranking after a private observation will be potentially blocking this influence. Given our assumptions about the naming function and the need for satisfying requirement 8, we must allow the observer to dictate the joint ranking which results from a private observation.

∎

References

[AGM85] C. Alchourron, P. Gardenfors, and D. Makinson. On the logic of theory change: Partial meet contraction functions and their associated revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[AK88] D. Appelt and K. Konolige. A Practical Nonmonotonic Theory for Reasoning about Speech Acts. In *Proc. Annual Meeting of the ACL*, pages 170–178, 1988.

[Arr63] K.J. Arrow. *Social Choice and Individual Values (2nd Ed.).* Yale University Press, 1963.

[Bou92] C. Boutilier. Normative, Subjective and Autoepistemic Defaults: Adpting the Ramsey Test. In *The 3rd international conference on principles of knowledge representation and reasoning*, pages 685–696, 1992.

[BT94] R. Brafman and M. Tennenholtz. Belief Ascription and Mental Level Modeling. In *Proc. of the 4th International Conference on Principles of Knowledge Representation and Reasoning*, pages 87–98, 1994.

[Doy91] J. Doyle. Rational Belief Revision. In J. Allen, R. Fikes, and E. Sandewall, editors, *Principles of Knowledge Representation and Reasoning*, pages 163–174. Morgan Kaufmann, 1991.

[dS92] A. delVal and Y. Shoham. Deriving Properties of Belief Update from Theories of Action. In *Proc. of AAAI-92*, pages 584–589, 1992.

[DW89] J. Doyle and M.P. Wellman. Impediments to Universal Preference-Based Default Theories. In *Proceedings of the 1st conference on principles of knowledge representation and reasoning*, 1989.

[FH94] N. Friedman and J.Y. Halpern. A Knowledge-Based Framework for Belief Change, Part II: Revision and Update. In *The 4th international conference on principles of knowledge representation and reasoning*, pages 190–201, 1994.

[Gal92] J. R. Galliers. Autonomous Belief Revision and Communication. In P. Gardenfors, editor, *Belief Revision*, pages 220–246. Cambridge University Press, 1992.

[Gar88] P. Gardenfors. *Knowledge is Flux.* The MIT Press, 1988.

[Gar92] P. Gardenfors. *Belief Revision.* Cambridge University Press, 1992.

[Gas91] G. Gaspar. Communication and Belief Changes in a Society of Agents: Towards a Formal Model of an Automanous Agent. In Y. Demazeau and J.P. Muller, editors, *Decentralized AI 2*, pages 245–255. North-Holland/Elsevier, 1991.

[GM88] P. Gardenfors and D. Makinson. Revision of Knowledge Systems using Epistemic Entrichment. In *The 2nd conference on Theoretical Aspects of Reasoning About Knowledge*, pages 83–95, 1988.

[Gro88] A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[GS90] B.J. Grosz and C.L. Sidner. Plans for Discourse. In P.R. Cohen, J. Morgan, and M.E. Pollack, editors, *Intentions in Communication*, pages 417–444. The MIT Press, 1990.

[HB91] M. N. Huhns and D. M. Bridgeland. Multiagent truth maintenance. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(6):1437–1445, 1991.

[KD96] Noa E. Kfir-Dahav. Multi-Agent Belief Revision. MSc. Thesis, Technion (in preparation), 1996.

[KM91] H. Katsuno and A. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294, 1991.

[LR57] R. D. Luce and H. Raiffa. *Games and Decisions- Introduction and Critical Survey.* John Wiley and Sons, 1957.

[Mor90] L. Morgenstern. A Formal Theory of Multiple Agent Nonmonotonic Reasoning. In *Proc. of AAAI-90*, pages 538–544, 1990.

[vdM94] R. van der Meyden. Mutual Belief Revision. In *The 4th international conference on principles of knowledge representation and reasoning*, pages 595–606, 1994.