(Dis)Belief Change and Feed-Back loop

Laurent Perrussel and Jean-Marc Thévenin
IRIT-Université Toulouse 1
Manufacture des Tabacs - 21 allée de Brienne
F-31042 Toulouse Cedex-France
{laurent.perrussel, jean-marc.thevenin}@univ-tlse1.fr

Abstract

This paper focuses on the features of belief change in a multi-agent context where agents consider beliefs and disbeliefs. Agents receive messages holding information from other agents and change their belief state accordingly. Agents may refuse to adopt incoming information if it is considered as not reliable. For this, agents maintain a preference relation over other agents embedded in the multi-agent system, in order to decide if they accept or reject the incoming information whenever inconsistencies occur. This process leads to non-prioritized belief revision. We focus here on this latter stage when agents refuse to change their (dis)beliefs and thus prefer to stay in opposite (dis)beliefs. In this case they inform their sources of information whether they actually changed their beliefs. We describe a process of justification where an agent states their preferences to the sender of the rejected information. This stage may lead the sender to also reconsider its own belief state.

Introduction

It is quite common to characterize intelligent agents in cognitive terms such as the well known belief, desire intention mental attitudes (Wooldridge & Jennings 1995). In that context, belief change is a key problem for agents. When a first agent sends a message holding information to a second agent, the first one intends to change the mental state of the second one. We consider two kinds of messages: messages about statements that do hold and thus entailing belief and messages about statements that do not hold and thus entailing disbeliefs (Ghose & Goebel 1998).
An agent should change its beliefs whenever it receives new information from other agents, i.e. it computes how its beliefs should look like after interpreting a message (Rao 1989; Perrussel 2003). In this paper, we consider non-prioritized belief change (Hansson 1999). In order to know if it has to adopt incoming information, the receiver considers the reliability of the sender. Whenever the incoming information introduces inconsistencies with information considered as more reliable, the receiver agent do not adopt it.

In this paper, we describe a process of justification, the feed-back loop, where the receiver motivates its reasons for ignoring incoming information by providing the sources of information inconsistent with it to the sender. If the sender considers the sources of the receiver as more reliable as its own sources then it also changes its beliefs (respectively disbeliefs).
We consider here that agents exchange messages about a world that do not change. In that context, belief change has to be considered as belief revision.

The paper is organized as follows: first we present an intuitive example. Next we introduce some formal definitions for describing messages and agent's belief states. Next, we present the definitions for the change functions. Finally, we describe how an agent change its beliefs after sending a message, according to the feed-back it receives about this message. We conclude the paper by discussing some open issues.

An intuitive example

Let us consider agent Peter (pe) receiving messages from agents Paul (pa) and the police department (po). Suppose that Peter considers the police department as more reliable as Paul ($pa <_{pe} po$). Paul tells to Peter that *John is a murderer* ($murd$). Peter adopts this statement and believes it. Paul also tells to Peter that *if John is a murderer, John will go to jail* ($murd \rightarrow jail$). The subset $\{murd, murd \rightarrow jail, jail\}$ is included in the belief set of Peter. Next, the police department tells to Peter that *without evidence John is not a murderer* ($murd \rightarrow evid$). The police tells to Peter to disbelieve there is evidence against John ($evid$). Because disbelief $evid$ is inconsistent with Peter's current beliefs, Peter change its beliefs. According to $pa <_{pe} po$, Peter no more believes that John is a murderer. Suppose now, that Paul tells once more to Peter that John is a murderer. Peter refuses to change its opinion about John and thus do not change its beliefs and disbeliefs. In fact disbelief *evidence* prevents Peter to adopt Paul's statement even if it is consistent with its belief. Peter informs Paul he does not adopt what Paul believes. Paul asks Peter to justify its refusal. Peter considers its belief and disbelief in contradiction with Paul's statement and informs Paul about agents supporting this inconsistency which he considers as more reliable. Consequently, if Paul considers Peter and its opinion about the police department as convincing, it also adopts the opinion of Peter and thus also changes its beliefs and disbeliefs.

Agent beliefs

We assume that beliefs are expressed in a propositional language $\mathcal{L}$. Changes in a belief set are caused by communication. We assume throughout the paper that the external world is *static*; handling changes caused by "physical" actions would require the integration of belief update to our formalism, which we leave for further work. Thus, we are considering cases such as diagnosis. We assume that messages are sent point-to-point. In order to identify the sender of messages we introduce a set of agent id: let $A = \{a, b \cdots\}$ be this set. We usually denote by s the sender agent and by r the receiver.

Describing messages

In our context, an agent may send two kinds of messages to other agents: agent s informs agent r that ϕ holds or agent s informs agent r that ϕ does not hold. We do not consider how agents acquire information. Agents interact with each others by sending and receiving messages. In more formal terms, we get:

Definition 1 (Message) *A message M is defined as a tuple of receiver r, sender s, content ϕ, status st. The receiver and the sender are agent ids, the content is an $\mathcal{L}$-formula and the status is one of the two possible status:* $\{\texttt{Hold}, \texttt{NotHold}\}$. *Self addressed messages are not allowed, i.e. $M = \langle r, s, \phi, st\rangle$ s.t. $s \neq r$ and $\phi \nvdash \bot$. Let $\mathcal{M}$ be the set of all possible messages.*

Notice that $\texttt{NotHold}\phi$ is not equivalent to $\texttt{Hold}\neg\phi$. Indeed $\texttt{NotHold}\phi$ means that ϕ should not be believed. Notice that agents may both disbelieve in a statement ϕ and its negation $\neg\phi$.

At each moment only one agent receives a message and change its state accordingly.

Definition 2 (Sequence of messages) *A sequence of messages σ is a function which associates moments in time and messages. Moments in time are represented by the set of integers: $\sigma : \mathbb{N} \rightarrow \mathcal{M}$.*

Describing agent beliefs

To represent agent's belief, we define a signed belief as a pair ⟨statement, origin of the statement⟩ (the sender of the statement):

Definition 3 (Signed belief) *A signed belief is a pair $\langle\phi, s\rangle$ where ϕ is a $\mathcal{L}$-formula and $s \in A$. Let $\mathcal{SB}$ be the set of all possible signed beliefs.*

Based on a set of signed beliefs, a belief set describes all its consequences:

Definition 4 (Belief set) *Let Bel be a function which maps a signed beliefs set $S \subseteq \mathcal{SB}$ to a set of $\mathcal{L}$-formulas: $Bel(S) = \{\psi | \bigwedge_{\langle\phi,s\rangle\in S} \phi \vdash \psi\}$. $Bel(S)$ represents the* belief set *associated to S.*

Example 1 *Let $S_{pe} = \{\langle murd, pa\rangle, \langle murd \rightarrow jail, pa\rangle\}$ be a signed belief set. The belief set associated to S_{pe} is: $Bel(S_{pe}) = \{murd, murd \rightarrow jail, jail...\}$.*

From a set of signed beliefs, we consider the minimal subsets entailing a specific conclusion. Let ϕ be a formula and S a set of signed beliefs. Let *support* be a function returning the set of minimal subsets of S entailing ϕ.

$$support(S, \phi) = \{S' | S' \subseteq S, \phi \in Bel(S') \text{ and } \forall S'' \subset S'(\phi \notin Bel(S''))\}$$

In order to describe what is actually believed by an agent, we introduce the notion of belief state. A belief state describes what is "currently" believed and what should not be believed by the agent.

- CB: a signed beliefs set representing current beliefs; This set changes with respect to the flow of messages about statements that do hold.
- DB: a signed beliefs set representing statements that should not be believed by the agent. The disbeliefs set changes with respect to the flow of messages about statements that do not hold.

Definition 5 (Belief state) *Let a be an agent. The belief state E_a of r is a pair $\langle CB_a, DB_a\rangle$ where $CB_a \subseteq \mathcal{SB}$ and $DB_a \subseteq \mathcal{SB}$. A belief state is defined with respect to a sequence of messages σ:*

- *$CB_a \subseteq \mathcal{SB}$, $(\forall\langle\phi, s\rangle \in CB_a)(\exists n)$ s.t. $(\sigma(n) = \langle r, s, \phi, \texttt{Hold}\rangle)$ and $Bel(CB_a) \nvdash \bot$;*
- *$DB_a \subseteq \mathcal{SB}$, $(\forall\langle\phi, s\rangle \in DB_a)((Bel(CB_a) \nvdash \phi)$ and $(\exists n)\sigma(n) = \langle r, s, \phi, \texttt{NotHold}\rangle)$.*

In other words, $Bel(CB_a)$ do not entail any disbelief.

Example 2 *Let us consider the intuitive example previously described. First $pa(ul)$ sends two messages to $pe(ter)$:*

$$\sigma = [\langle pe, pa, murd, \texttt{Hold}\rangle, \langle pe, pa, murd \rightarrow jail, \texttt{Hold}\rangle]$$

Suppose peter "adopts" the messages, its belief state E_{pe} looks like:

$$\langle\{\langle murd, pa\rangle, \langle murd \rightarrow jail, pa\rangle\}, \emptyset\rangle$$

E_{pe} satisfies the constraints mentioned definition 5. Next peter processes the messages sent by the police (if john is a murderer then there is evidence against him; and the police do not believe there is evidence).

$$\begin{aligned} \sigma(3) &= \langle pe, po, murd \rightarrow evid, \texttt{Hold}\rangle \\ \sigma(4) &= \langle pe, po, evid, \texttt{NotHold}\rangle \end{aligned}$$

The subsequent belief state is

$$\langle\{\langle murd, pa\rangle, \langle murd \rightarrow jail, pa\rangle, \langle murd \rightarrow evid, po\rangle\}, \{\langle evid, po\rangle\}\rangle$$

and is not valid since it violates the constraints about disbelief $evid(ence)$ entailed by current beliefs. Thus, $pe(ter)$

has to revise its set CB_{pe} in order to adopt disbelief $evid$. It may remove, for instance, the pair $\langle murd, pa\rangle$ from its current beliefs in order to get a consistent belief state. Thus new belief state E_{pe} defined as following is well-defined:

$$\langle\{\langle murd \rightarrow jail, pa\rangle, \langle murd \rightarrow evid, po\rangle\}, \{\langle evid, po\rangle\}\rangle$$

Belief state change

Each agent a uses a procedure for changing its belief state whenever it receives new messages, in order to decide which signed beliefs should belong to CB_a and DB_a. This procedure requires that each agent can produce an entrenchment ordering over its signed beliefs. This ordering is based on the reliability of agents. Each agent a represents the reliability of agents with an order over agents $\leqslant_a$. Agents that could not be distinguished are considered in an equal way (which entails a total preorder). Writing $b <_a c$ means that c is a strictly better source than b: $b \leqslant_a c$ but $c \nleqslant_a b$.
Based on total preorder $\leqslant_a$, agent a can produce a total preorder over its signed beliefs. In this paper, if $b \leqslant_a c$ then for all ϕ, ψ $\langle\phi, b\rangle \leqslant_a \langle\psi, c\rangle$. Notice that it is possible for agents to produce more sophisticated entrenchment ordering over their signed beliefs. For instance they could partition the vocabulary and define specific preorders over agents for each partition (del Cerro *et al.* 1998).
In our context, we consider non-prioritized versions of belief change (Booth 2002). Indeed, according to their entrenchment ordering, agent may refuse the incoming statements. In addition, the belief change functions have to consider disbeliefs. We describe below the change functions $*$ and $-$ using standard change functions (Gärdenfors 1988) named *cbr* and *cbc*; the aim of *cbr* and *cbc* functions is to change CB sets in a prioritized way while the aim of $*$ and $-$ is to change belief states in a non-prioritized way. In the revision literature (Gärdenfors 1988), functions such as *cbr* and *cbc* have been mainly discussed to perform respectively revision and prioritized contraction on current beliefs. We consider that both functions *cbr* and *cbc* satisfied at least the success AGM postulate (Gärdenfors 1988). To enforce agents autonomy, each agent has its own functions *cbr* and *cbc* based on its own preferences.

Definition 6 (Agent revision program) *Let p_a be the revision program of agent a defined as a structure: $\langle\leqslant_a, cbr_a, cbc_a\rangle$. $\leqslant_a$ is a total preorder over A and cbr_a and cbc_a are agent a's functions describing, respectively, the revision and contraction of its current beliefs with respect to $\leqslant_a$: cbr_a: $2^{\mathcal{SB}} \times \mathcal{SB} \times 2^{A\times A} \rightarrow 2^{\mathcal{SB}}$, cbc_a: $2^{\mathcal{SB}} \times \mathcal{SB} \times 2^{A\times A} \rightarrow 2^{\mathcal{SB}}$. Let $\mathcal{P}$ be the set of agent revision programs.*

The revision program is appropriate for handling iterated belief change (Darwiche & Pearl 1997) since the selection mechanism is only based on $\leqslant_a$ and not specific to a belief state.

In the following, we show how pair $E_r = \langle CB_r, DB_r\rangle$ changes when agents receive messages. Let us consider a set of agents A where their initial belief states is empty: $(\forall a \in A) E_a^0 = \langle\emptyset, \emptyset\rangle$, a set of revision programs p_a and a sequence of messages σ. Messages received by agents entail a revision action, denoted by $*$, or a contraction action, denoted by $-$:

Definition 7 (Changing receiver belief state) *Let a be any agent, E_a^{n-1} its belief state at $n-1$ and $\sigma(n) = \langle r, s, \phi, status\rangle$ be a message.*

$$E_a^n = \begin{cases} (E_a^{n-1}) & \text{if } a \neq r \text{ and } a \neq s \\ (E_a^{n-1})^*_{\langle\phi,s\rangle} & \text{if } status = \texttt{Hold} \text{ and } a = r \\ (E_a^{n-1})^-_{\langle\phi,s\rangle} & \text{if } status = \texttt{NotHold} \text{ and } a = r \end{cases}$$

The change actions for belief state of s will be given next section.

Notice that functions *cbr* and *cbc* are wrapped into functions $*$ and $-$ as defined below. Agents are autonomous: for each message they receive, they use their own revision program in order to prevent inconsistencies in their beliefs. The belief state is recursively defined accordingly to p_r for any message received at a moment $m < n$ by r.

Revision of belief state

When the belief state of agent r is revised, we distinguish two cases:

- signed belief $\langle\phi, s\rangle$ is more reliable, w.r.t. $\leqslant_r$ than conflicting signed beliefs and $\langle\phi, s\rangle$ is also more reliable than all conflicting disbeliefs: agent s has to be trusted and ϕ has to believed by r;
- otherwise: $\langle\phi, s\rangle$ is not sufficiently reliable and r ignores statement ϕ.

According to this framework, we first specify the condition stating if agent r has to revise its belief state and, second, we describe function $*$.

Definition 8 (R_Ignore) *Let E_r be the belief state of agent r and $p_r = \langle\leqslant_r, cbr_r, cbc_r\rangle$ its revision program; let $\langle\phi, s\rangle$ be a signed belief. The condition* $\mathbf{R\_Ignore}$ *stating if agent r should not revise its belief state is defined as follows:*

$$(\exists\langle\psi, a\rangle \in DB_r \text{ s.t. } \psi \in Bel(cbr_r(CB_r, \langle\phi, s\rangle, \leqslant_r)) \text{ and }$$
$$\forall\gamma \in support(cbr_r(CB_r, \langle\phi, s\rangle, \leqslant_r), \psi) \max(\gamma) <_r \langle\psi, a\rangle)$$
$$\text{or} (\neg\phi \in Bel(CB_r) \text{ and } \exists\gamma \in support(CB_r, \neg\phi)$$
$$\text{s.t. } \langle\phi, s\rangle <_r \min(\gamma))$$

First part of the disjunction specifies "if there is any reliable disbeliefs that will be violated" if the revision occurs. Second part checks whether the negation of the input is already in the belief state and checks its reliability.

Definition 9 ($*$) *Let $n \in \mathbb{N}$, $M = \sigma(n) = \langle r, s, \phi, st\rangle$ be a message such that $st = \texttt{Hold}$, and $E_r^{n-1} = \langle CB_r, DB_r\rangle$ be the belief state of agent r at $n-1$. Let $(E_r^{n-1})^*_{\langle\phi,s\rangle}$ be the resulting belief state of r after revising E_r^{n-1} by $\langle\phi, s\rangle$:*

- *if condition* $\mathbf{R\_Ignore}$ *holds then* $(E_r^{n-1})^*_{\langle\phi,s\rangle} = \langle CB_r, DB_r\rangle$
- *else* $(E_r^{n-1})^*_{\langle\phi,s\rangle} = \langle cbr_r(CB_r, \langle\phi,s\rangle, \leqslant_r), DB_r - \Delta_r\rangle$

Disbeliefs that should no longer be considered have to be removed from DB. Let Δ_r be this set:

$$\Delta_r = \{\langle\psi,a\rangle \in DB_r | \psi \in Bel(cbr_r(CB_r, \langle\phi,s\rangle, \leqslant_r)) \text{ and } (\exists\gamma \in support(cbr_r(CB_r, \langle\phi,s\rangle, \leqslant_r), \psi)\ \langle\psi,a\rangle <_r \max(\gamma))\}$$

Contraction of belief state

In this section, we briefly describe the contraction operator $-$. This action is very close to the revision operator $*$ but could not be defined by considering $*$ and the Harper identity. Thus, $*$ and $-$ are two separate actions. At first, current beliefs set is contracted if the sender is reliable and second, the input statement is added to the set of disbeliefs. As previously, we first specify a condition stating if an agent has to contract its epistemic state and second we describe function $-$.

Definition 10 (C_Ignore) *The condition* $\mathbf{C\_Ignore}$ *stating if an agent should not contract its belief state is defined as follows:*

$$\phi \in Bel(CB_r) \text{ and } \exists\gamma \in support(CB_r, \phi) \text{ s.t. } \langle\phi,s\rangle <_r \min(\gamma)$$

This condition checks whether ϕ already belongs to the belief set and if agents supporting ϕ are more reliable than s.

Definition 11 (-) *Let $(E_r^{n-1})^-_{\langle\phi,n\rangle}$ be the belief state of r at n after contracting E_r^{n-1} by $\langle\phi,n\rangle$:*

- *if condition* $\mathbf{C\_Ignore}$ *then* $(E_r^{n-1})^-_{\langle\phi,n\rangle} = \langle CB_r, DB_r\rangle$
- *else* $(E_r^{n-1})^-_{\langle\phi,n\rangle} = \langle cbc_r(CB_r, \langle\phi,s\rangle, \leqslant_r), DB_r \cup \{\langle\phi,s\rangle\}\rangle$

Example 3 *Let us consider again example 2. Preferences of peter are: $\{pa <_{pe} pe, pe <_{pe} po\}$. According to example 2, peter has the following belief state E_{pe}^4 at moment 4:*

$$\langle\{\langle murd \rightarrow jail, pa\rangle, \langle murd \rightarrow evid, po\rangle\}, \{\langle evid, po\rangle\}\rangle$$

In other words, signed belief $\langle murd, pa\rangle$ has been removed of the current beliefs set CB_{pe}^3 by function cbc_{pe}. At this moment, peter does no longer believe that John is a murderer and he will go to jail. Next paul informs again peter that john is a murderer. We have:

$$\sigma(5) = \langle pe, pa, murd, \texttt{Hold}\rangle$$

Even if the statement $murd$ is consistent with current belief of peter, peter ignores the message of paul. Condition $\mathbf{R\_Ignore}$ *checks that (i) disbelief $evid$ would be violated since $murd$ entails $evid$ and (ii) this disbelief is issued from an agent that is more reliable than paul. Thus peter's belief state is unchanged: $E_{pe}^5 = E_{pe}^4$.*

Feed-Back about changes

After changing its beliefs, agent r inform s if it has adopted or not the received information ϕ. By considering a feedback on the message, s may refine its own belief state. It seems relevant that whenever r replies that it actually adopts the received information, the dialog is closed. In the other case, s will re-consider its belief state. We do not consider the case where r or s propagate their change to other agents . This choice is mainly motivated by the fact that we can reach unsteady states where agents do not stop to change their belief states. Let us stress that the proposed feedback framework is common to every agent. However, we could consider that every agent has its own feedback protocol described in its revision program. We leave for further work this open problem. In the following we propose one protocol linking revision and contraction actions.

Now, let us focus on the case where s re-considers its belief state. Agent s behaves as follows with the objective to improve its belief state. It requests to r a justification which consists of a set of agent ids involved in agent r's signed beliefs (and disbeliefs) in contradiction with ϕ. From agent r's point of view ($\leqslant_r$), these agents are more reliable than s. From the sender's viewpoint, the alternative is:

- some of these agents in the justification are considered more reliable than agents supporting ϕ, according to $\leqslant_s$, and s has to change its mental attitude about ϕ.
- None of these agents are more reliable than the agents supporting ϕ according to $\leqslant_s$ and thus s pursues its mental attitude about ϕ.

Let us extend our framework for considering this proposal.

Feed-back concerning a revision action

Here we are considering the case where r has received a message from s about ϕ with status `Hold` and according to condition $\mathbf{R\_Ignore}$ r has not changed its belief state. According to the protocol, the justification stage consists of extracting the set of agent ids involved in the support of formulas that have led r to refuse the message of s. This set of agent ids is partitioned in two sets R_B and R_D.

Set R_B considers signed beliefs of r which prevented the revision action:

$$a \in R_B(n) \iff \forall\gamma \in support(CB_r^{n-1}, \neg\phi)\langle\psi,a\rangle \in \gamma$$

Set R_D considers disbelief of r:

$$a \in R_D(n) \iff \langle\psi,a\rangle \in DB_r^{n-1} \text{ and } Bel(cbr_r(CB_r^{n-1}, \langle\phi,s\rangle, \leqslant_r)) \vdash \psi \text{ and } \forall\gamma \in support(cbr_r(CB_r^{n-1}, \langle\phi,s\rangle, \leqslant_r), \psi) \max(\gamma) < \langle\psi,a\rangle$$

Agent s evaluates agents in R_B and R_D with respect to $\leqslant_s$ and decides if it pursues its mental attitude. Indeed, s considers sources of its own set of signed beliefs involved in the support of ϕ and compares them to the sets R_B and R_D. If the latters are better than the former then s changes its mental attitude, i.e. agent s changes its belief state. Let $R_\phi^{\leqslant}$

be a set of agent ids representing the more reliable agents involved in the support of current belief ϕ:

$$R^{\leqslant}_{\phi}(n) = \{a|\langle\psi, a\rangle \in \bigcup_{S \in support(CB^{n-1}_s,\phi)} \max(S)\}$$

Since s may change its belief state, we extend the change function described in def. 7 by considering the sender's belief state. Now, a message may trigger a change action on the sender's belief state:

Definition 12 (Changing sender belief state) *Let E^{n-1}_s be the belief state of agent s, $\sigma(n) = \langle r, s, \phi, \mathtt{Hold}\rangle$ be a message. E^n_s is defined a s follows:*

1. *If condition* $\mathbf{R\_Ignore}$ *does not hold then $E^n_s = E^{n-1}_s$.*
2. *Otherwise:*
   (a) *if $(\forall a \in R_B(n) \cup R_D(n))(\exists b \in R^{\leqslant}_{\phi}(n))\ a <_s b$ then $E^n_s = E^{n-1}_s$.*
   (b) *if $(\exists a \in R_B(n))(\forall b \in R^{\leqslant}_{\phi}(n))\ a \not<_s b$ then $E^n_s = (E^{n-1}_s)^*_{\langle\neg\phi,a\rangle}$.*
   (c) *if $(\exists a \in R_D(n))(\forall b \in R^{\leqslant}_{\phi}(n))\ a \not<_s b$ then $E^n_s = (E^{n-1}_s)^-_{\langle\phi,a\rangle}$.*

Notice the difference for the change operations between case (b) and case (c).

- Case (b): the change action is a revision action since $a \in R_B$ which means that $\neg\phi$ should be believed by s thanks to a.

- Case (c): the change action is a contraction action since $a \in R_D$; future disbelief ϕ has to be supported by agents issued from R_D.

The main consequence of the feed-back is a reinforcement of (dis)beliefs of s. This is due to the fact that s changes its belief state only if r justifies its refusal by exhibiting better sources of (dis)belief:

Proposition 1 *Let E^n_s be the belief state of s at n.*

$$\forall a(a \in R^{\leqslant}_{\phi}(n)\ \exists b \in R^{\leqslant}_{\phi}(n+1) \cup R^{\leqslant}_{\neg\phi}(n+1) \Rightarrow a \leqslant_s b)$$
$$\text{or } \forall a \in R^{\leqslant}_{\phi}(n)(\exists b(\langle\phi, b\rangle \in DB^n_s \Rightarrow a \leqslant_s b))$$

First part of the disjunction states that for each agent a which was one of the less reliable involved in the support of ϕ, there exists an agent b supporting ϕ or $\neg\phi$ in the new belief state s.t. $a \leqslant_s b$ (improvement of the belief sources). Second part states that agents b supporting disbelief ϕ are at least as reliable as all agents a which were the less reliable supports of belief ϕ (improvement of (dis)belief sources).

A second consequence of the feed-back loop is that change actions are not only caused by messages but also by the evaluation of these change actions.

Example 4 *Let us pursue example 3 and focus on the last message. After handling the sequence of messages $\sigma(1)$ to $\sigma(5)$ (see above), we have the following belief state for peter:*

$$\langle\{\langle murd \rightarrow jail, pa\rangle, \langle murd \rightarrow evid, po\rangle\}, \{\langle evid, po\rangle\}\rangle$$

Since condition $\mathbf{R\_Ignore}$ *holds, peter do not adopt message $\sigma(5)$. Suppose that preferences of paul are: $\{(pe < pa, pa < po)\}$ and its belief state is $\langle\{\langle murd, pa\rangle\}, \emptyset\rangle$. We get $R_B = \emptyset$ and $R_D = \{po\}$. As paul considers agent police as a reliable agent, paul contracts its belief state by $\langle murd, po\rangle$ (case (c) of def. 12).*

Feed-back concerning a contraction action

Here we are considering the case where r has received a message from s about ϕ with status `NotHold` and according to condition $\mathbf{C\_Ignore}$ r has not changed its belief state. The feed-back loop for a contraction action is close to the feed-back loop for revision action: Let C_B bet a set of agent ids representing origin of beliefs of r which prevents the contraction action:

$$a \in C_B(n) \iff \exists\gamma \in support(CB^{n-1}_r, \phi) \text{ s.t. } \langle\psi, a\rangle \in \min(\gamma) \text{ and } \langle\phi, s\rangle <_r \langle\psi, a\rangle$$

When we consider feed-back about contraction, it is useless to consider disbeliefs since no disbelief can be violated. Next, agent s evaluates agents in C_B with respect to $\leqslant_s$ and decides if it pursues its mental attitude. Let $C^{\leqslant}_{\phi}$ be a set of agent ids representing agents involved in disbelief ϕ:

$$C^{\leqslant}_{\phi}(n) = \{a|\langle\phi, a\rangle \in DB^{n-1}_s\}$$

We extend the change function described in def. 12 by considering the sender's belief state change. We consider two cases: s has to revise its current beliefs or not.

Definition 13 (Changing sender belief state) *Let E^{n-1}_s be the belief state of agent s, $\sigma(n) = \langle r, s, \phi, \mathtt{NotHold}\rangle$ be a message. E^n_s is defined a s follows:*

1. *If condition* $\mathbf{C\_Ignore}$ *does not hold then $E^n_s = E^{n-1}_s$.*
2. *Otherwise:*
   (a) *if $(\forall a \in C_B(n))(\exists b \in C^{\leqslant}_{\phi}(n))\ a <_s b$ then $E^n_s = E^{n-1}_s$.*
   (b) *if $(\exists a \in C_B(n))(\forall b \in C^{\leqslant}_{\phi}(n))\ a \not<_s b$ then $E^n_s = (E^{n-1}_s)^*_{\langle\phi,a\rangle}$.*

As we can see, a contraction action ignored at the receiver level may entail a revision action at the sender level. In case (a), s keeps its opinion about ϕ since the justification of r is not relevant. In case (b), the justification provided by r is relevant: some agents a belonging to $C_B(n)$ are considered by s as reliable and thus s revises its belief state by $\langle\phi, a\rangle$.

As previously, feed-back loop entails a reinforcement of beliefs of s:

Proposition 2 *Let E_s^n be the belief state of s at n.*

$$\forall a \in C_\phi^{\leqslant}(n) \text{ if } \exists b \in R_\phi^{\leqslant}(n+1) \text{ then } a \leqslant_s b)$$

In other words, if s has actually replaced disbelief ϕ by belief ϕ then the reliability of belief ϕ is at least as good as the reliability of previous disbelief ϕ.

Conclusion

In this paper, we have presented belief change operators with a feed-back loop. Our work focus on non-prioritized iterated belief change when agents exchange information about a static world such as diagnosis. The first characteristic of our proposal is to define change by considering revision and contraction as two independant actions by considering disbelief in an explicit way and also by introducing a feed-back loop. This loop shows that a revision action performed by the receiver of a message may entail a contraction action for the sender of the message. The second characteristic of our proposal is that messages not only improve beliefs, respectively disbeliefs, in the receiver's belief state but also in the sender's belief state (i.e. mutual improvement).

Our approach does have some limitations. In this paper, we only look at cases where belief change concerns a static world. If we consider more general actions, changes may result from sensing actions and these changes may be propagated to other agents through communicative actions. We would like to generalize our framework in order to handle these limitations, i.e. update rather than revision. A second extension concerns the feed-back loop: a first issue is to define a family of protocols rather than one protocol. For instance, agents may justify their refusal by providing the previous messages that led them to this refusal. This allows the sender to get a piece of story that it has missed. A second issue is to specify our justification protocol in a speech act framework. Using speech act operators agents can send richer messages to assert statements, refuse statements and justify their attitudes. We are currently investigating these topics.

References

Booth, R. 2002. the logic of iterated non-prioritised revision. In *Proceedings of the Workshop on Conditionals, Information and Inference. Hagen, Germany*.

Darwiche, A., and Pearl, J. 1997. On the logic of Iterated Belief Revision. *Artificial intelligence* 89(1):1–29.

del Cerro, L. F.; Herzig, A.; Longin, D.; and Rifi, O. 1998. Belief reconstruction in cooperative dialogues. In Giunchiglia, F., ed., *Proc. 8th Int. Conf. on Artificial Intelligence: Methods, Systems, Applications (AIMSA'98)*, volume 1480 of *Lecture Notes in Computer Science*, 254–266. Springer Verlag.

Gärdenfors, P. 1988. *Knowledge in flux: Modeling the Dynamics of Epistemic States*. MIT Press.

Ghose, A., and Goebel, R. 1998. Belief states as default theories: studies in non-prioritised belief change. In Prade, H., ed., *Proceedings of the 13th European Conference on Artificial Intelligence–ECAI'98*, 8–12.

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

Perrussel, L. 2003. Handling sequences of belief change in a multi-agent context. In Schillo, M.; Klusch, M.; Muller, J.; and Tianfield, H., eds., *Multiagent System Technologies. First German Conference, MATES 2003, Erfurt, Germany, September 22-25, 2003, Proceedings*, volume 2831. Springer.

Rao, A. 1989. Dynamics of belief systems: A philosophical, logical, and ai perspective. Technical Report 02, Australian Artificial Intelligence Institute, Melbourne, Australia.

Wooldridge, M., and Jennings, N. R. 1995. Intelligent agents: Theory and practice. *The Knowledge Engineering Review* 10(2):115–152.