ECAI 2006
G. Brewka et al. (Eds.)
IOS Press, 2006

Mutual Enrichment for Agents Through Nested Belief Change: A Semantic Approach

Laurent Perrussel and **Jean-Marc Thévenin**[1] and **Thomas Meyer**[2]

1 Introduction

This paper focuses on the dynamics of nested beliefs in the context of agent interactions. Nested beliefs represent what agents believe about the beliefs of other agents. We consider the *tell* KQML performative [1] which allows agents to send their own beliefs to others. Whenever agents accept a new belief, or refuse to change their own beliefs after receiving a message, both receiver and sender enrich their nested beliefs by refining their beliefs about (i) the other agent's beliefs and (ii) the preferences of the other agent. The main objective of nested beliefs is to improve cooperation between agents. We propose a logical formalisation of the acquisition process of nested beliefs and preferences. This acquisition process is the first step towards the elaboration of sophisticated interaction protocols.

2 The Logical Framework

To represent an agent's beliefs we use signed statements. A *signed statement* is a pair ⟨statement, origin of the statement⟩ (usually the sender of the statement). Let $\mathcal{L}_0$ be a propositional language and A a set of agent identities. We define a signed statement as a pair $\langle \phi_0, a\rangle$ where ϕ_0 is a $\mathcal{L}_0$-formula and $a \in A$ is the origin of ϕ_0. Let $\mathcal{S}$ be the set of all sets of signed statements: $\mathcal{S} = 2^{\mathcal{L}_0 \times A}$. The *belief state* of an agent is a pair ⟨set of signed statements, set of sets of signed statements⟩. The first set describes the *basic beliefs* of the agent: what it currently believes. The second set describes the *nested beliefs* of the agent: what it believes about the basic beliefs of others.

Definition 1 (Belief state) *A belief state BS_a^n of agent a is a pair $\langle CB_a^n, NB_a^n\rangle$ s.t. (i) $CB_a^n \in \mathcal{S}$ represents the basic beliefs of agent a at time n and (ii) $\forall CB_{a,b}^n \in NB_a^n, CB_{a,b}^n \in \mathcal{S}$ represents the nested beliefs of agent a about agent b at n. Let $\mathcal{B}$ be the set of all possible belief states.*

Agents revise their basic beliefs and nested beliefs each time they receive a *tell* performative. Let S be a set of signed beliefs and $*$ be a revision operator [2]; $S^*_{\langle \phi_0, a\rangle}$ denotes the revision of S by $\langle \phi_0, a\rangle$.

Preferences may be defined taking various matters into account [5]. We simply assume that agents have preferences over the set of agents A which describe the reliability of the sources of information, i.e. the level of trust an agent has about the other agents with which it interacts. We suppose that agents are equally reliable when they can't be distinguished, which entails a total preorder. Let $\preceq_a^n$ be a total preorder over A representing agent a's preferences at time n. $b \preceq_a^n c$ stands for agent c is at least as preferred as b for agent a at time n. As is the case for beliefs, agents can handle nested preferences. Nested preferences represent what agents believe about the preferences of other agents. $c \preceq_{a,b}^n d$ means: agent a believes that for agent b agent d is at least as preferred as c at n.

Definition 2 (Preference state) *A preference state PS_a^n of agent a is a pair $\langle \preceq_a^n, NP_a^n\rangle$ s.t. (i) $\preceq_a^n$ is a total preorder representing basic preferences of agent a at time n and (ii) every $\preceq_{a,b}^n \in NP_a$ is a total preorder representing agent b's preferences according to a at time n. $\mathcal{P}$ is the set of all possible preference states.*

Let $\langle BS_a^n, PS_a^n\rangle = \langle\langle CB_a^n, NB_a^n\rangle, \langle \preceq_a^n, NP_a^n\rangle\rangle$ be a tuple which represents the whole state of agent a at n, i.e. all its beliefs and preferences.

Let S be any set of signed statements (basic or nested beliefs) and $\preceq$ the corresponding preference relation (basic or nested). The logical closure of S w.r.t. $\preceq$ is obtained as follows. By $\min(S, \preceq)$ we denote the set of the least preferred agent identities w.r.t. $\preceq$ among agent identities signing beliefs of S. We suppose that statements entailed by S are signed with the least preferred agent identities of the minimal subsets of S entailing them: $\mathrm{Cn}(S, \preceq) = \{\langle \psi_0, a\rangle | \exists S' \subseteq S$ s.t. $\bigwedge_{\langle \phi_0, b\rangle \in S'} \phi_0 \models_{\mathcal{L}_0} \psi_0$ and $\nexists S'' \subset S'$ s.t. $\bigwedge_{\langle \phi_0, b\rangle \in S''} \phi_0 \models_{\mathcal{L}_0} \psi_0$ and $a \in \min(S', \preceq)\} \cup \{\langle \psi_0, a\rangle | \models_{\mathcal{L}_0} \psi_0$ and $a \in A\}$.

Now, we present the *action performatives* which lead to the dynamics of belief and preference states change. When an agent issues a *tell* performative to inform a receiver agent about its basic beliefs the receiver uses a prioritised belief revision operator $*$ to change its nested beliefs about the sender. As an acknowledgment of the *tell* performative, the receiver informs the sender with an *accept* (respectively *deny*) performative if the incoming statement has been incorporated in its basic beliefs. The sender in turn applies prioritised revision to its nested beliefs about the receiver. More formally:

- $\texttt{Tell}(s, r, \phi_0, a)$ stands for: agent s informs r that it believes ϕ_0 signed by a according to the standard KQML semantics [1]. When receiving a `Tell` performative, agent r revises its belief state by $\langle \phi_0, a\rangle$ in a non-prioritised way [3] according to its preferences.
- $\texttt{Accept}(r, s, p)$ stands for: agent r informs s that it accepts the performative p. If $p = \texttt{Tell}(s, r, \phi_0, a)$ then $\texttt{Accept}(r, s, p)$ means that agent r has revised its basic beliefs by $\langle \phi_0, a\rangle$ and thus believes ϕ_0.
- $\texttt{Deny}(r, s, p)$ stands for: agent r informs s that it refuses to process

[1] IRIT–Université Toulouse 1, Manufacture des Tabacs, 21 allée de Brienne, F-31042, Toulouse Cedex - France, email: laurent.perrussel@univ-tlse1.fr, email: jean-marc.thevenin@univ-tlse1.fr

[2] National ICT Australia and University of New South Wales, Sydney, Australia, email: thomas.meyer@nicta.com.au. National ICT Australia is funded by the Australia Government's Department of Communications, Information and Technology and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Centre of Excellence program. It is supported by its members the Australian National University, University of NSW, ACT Government, NSW Government and affiliate partner University of Sydney.

the performative p. If performative $p = \texttt{Tell}(s, r, \phi_0, a)$ it means that agent r has not revised its basic beliefs by $\langle \phi_0, a \rangle$.

The dynamics of the system is given by the execution of performatives. A sequence of actions σ is a function which associates integers representing state labels with performatives.

3 The Dynamics of Mutual Enrichment

Let us first express the honesty postulate (Hon) as follows. This postulate, which is recommended in a cooperative context, states that if a `Tell` performative occurs, then at the same time the sender believes the corresponding signed statement. This actually enforces the standard KQML semantics of `Tell` [1].

(Hon) For any n if $\sigma(n) = \texttt{Tell}(s, r, \phi_0, a)$ then $\langle \phi_0, a \rangle \in \mathrm{Cn}(CB^n_s, \preceq^n_s)$.

Basically, after a performative $\texttt{Tell}(s, r, \phi_0, a)$, the belief state and preference state of agents do not change, except for the receiver. The receiver applies a non-prioritised revision of its basic beliefs using its basic preferences. It also applies a prioritised revision of its nested beliefs about the sender since it believes ϕ_0 signed by a according to the honesty postulate. It can also refine its nested preferences about the sender if it has removed a nested belief $\neg\phi_0$ signed by an agent b during the prioritised revision. Indeed the sender has preferred a to b. Finally the receiver acknowledges the `Tell` performative with an `Accept` or `Deny` performative, as formalized below, depending on the outcome of the non-prioritised revision of its basic beliefs so that the sender can in turn change its nested beliefs and nested preferences about the receiver. The non-prioritised revision is based on basic preferences in the following way: if the receiver believes $\neg\phi_0$ and the signatures of $\neg\phi_0$ are at least as preferred as a (the signature of ϕ_0 in the `Tell`), the receiver denies the `Tell` performative.

(DT) $\sigma(n+1) = \texttt{Deny}(r, s, \texttt{Tell}(s, r, \phi_0, a))$ iff $\sigma(n) = \texttt{Tell}(s, r, \phi_0, a)$ $\&\exists \langle \neg\phi_0, b \rangle \in \mathrm{Cn}(CB^n_r, \preceq_r)$ s.t. $a \prec_r b$

Otherwise there is no conflict or a is strictly more preferred than all the signatures of $\neg\phi_0$, and r thus accepts the `Tell` performative.

(AT) $\sigma(n+1) = \texttt{Accept}(r, s, \texttt{Tell}(s, r, \phi_0, a))$ iff $\sigma(n) = \texttt{Tell}(s, r, \phi_0, a)$ $\&\forall \langle \neg\phi_0, b \rangle \in \mathrm{Cn}(CB^n_r, \preceq_r), b \prec_r a$

Due to space restrictions we focus on the formulas describing how nested beliefs and nested preferences of the sender change (see [6] for the complete formalisation of the dynamics).

According to (AT), if r accepts the message of s, agent s believes that r believes ϕ_0 and thus does not believe $\neg\phi_0$, which results in s revising its nested beliefs about r with $\langle \phi_0, a \rangle$ (a prioritised revision).

(AdNB3-1) If $\sigma(n) = \texttt{Accept}(r, s, \texttt{Tell}(s, r, \phi_0, a))$ then $CB^{n+1}_{s,r} = (CB^n_{s,r})^*_{\langle \phi_0, a \rangle}$.

According to (DT), if r refuses the `Tell` performative then s concludes that r believes $\neg\phi_0$ and r believes that the signature of $\neg\phi_0$ has to be more trusted than a. So s revises its nested beliefs about r with the signed statement $\langle \neg\phi_0, b \rangle$ (again, a prioritised revision) so that signature b of $\neg\phi_0$ is preferred to a w.r.t. the nested preferences of s about r:

(AdNB3-2) If $\sigma(n) = \texttt{Deny}(r, s, \texttt{Tell}(s, r, \phi_0, a))$ then $CB^{n+1}_{s,r} = (CB^n_{s,r})^*_{\langle \neg\phi_0, b \rangle}$ s.t. $a \preceq_{s,r} b$

This bring us to the sender's nested preferences. Firstly, the sender's nested preferences about agents other than r do not change. Next, agent s draws conclusions about a depending on whether r accepts or denies the performative $\texttt{Tell}(s, r, \phi_0, a)$. Whenever agent r accepts, agent s refines its nested preferences only if s currently believes that r believes $\neg\phi_0$ so that a be strictly more preferred than the signatures of $\neg\phi_0$ If r denies the message, the nested preferences of s about the signatures of $\neg\phi_0$ change. According to condition (DT) they have to be as preferred as a. Finally, we need to take care of nested preferences, about r, of s that do not change. In order to formalise these requirements, we give an explicit procedure to change the nested preferences. The `Accept` performative helps the sender to refine its nested preferences since it allows to remove some preferences; i.e. it helps agent s to go toward a stricter order. Removing preferences means that agent s already believes that r believes $\neg\phi_0$ and thus a is strictly preferred to all the signatures of $\neg\phi_0$.

(Ad-KeNP2) Let $\langle CB^n_{s,r}, \preceq^n_{s,r} \rangle$ be the nested beliefs and preferences of s about r. Let $\sigma(n) = \texttt{Tell}(s, r, \phi_0, a)$ and $\sigma(n+1) = \texttt{Accept}(r, s, \texttt{Tell}(s, r, \phi_0, a))$. All nested preferences at n are propagated at time $n+1$ as follows: $\preceq^{n+1}_{s,r} = \preceq^n_{s,r} - \{a \preceq^n_{s,r} b | \langle \neg\phi_0, b \rangle \in CB^n_{s,r}\} \cup \{b \preceq^n_{s,r} a | \langle \neg\phi_0, b \rangle \in CB^n_{s,r}\}$.

If r does not accept the incoming message, agent s also changes its nested beliefs about r. The `Deny` performative does not help agent s to refine its nested preferences since we only add nested preferences.

(AdNP3) Let $\langle CB^n_{s,r}, \preceq^n_{s,r} \rangle$ be the nested beliefs and preferences of s about r. Let $\sigma(n) = \texttt{Tell}(s, r, \phi_0, a)$ and $\sigma(n+1) = Deny(r, s, \texttt{Tell}(s, r, \phi_0, a))$. All nested preferences at n are propagated at time $n+1$ as follows: $\preceq^{n+1}_{s,r} = \preceq^n_{s,r} \cup \{a \preceq^n_{s,r} b | \langle \neg\phi_0, b \rangle \in CB^n_{s,r}\}$.

It is easily shown that the conditions preserves the ordering for nested preference as a total preorder.

4 Conclusion

In this paper we have given a sketch of a formalisation for handling the dynamics of nested beliefs and preferences in the context of agent interactions where agents are cooperative (see [6] for the detailed formalisation). We have shown how agents acquire nested beliefs and preferences. For this we have presented a logical framework to describe nested beliefs, preferences, and performatives. This framework is useful for specifying properly the expected behaviour of agents handling the `Tell`, `Accept` and `Deny` performatives. We have started to build a logical language based on dynamic epistemic logic [4] in order to reason about dialogues. Our aim is to define a semantics based on the semantics proposed in this paper.

REFERENCES

[1] T. Finin, Y. Labrou, and J. Mayfield, 'KQML as an agent communication language', in *Software Agents*, ed., J. Bradshaw, MIT Press, (1997).
[2] P. Gärdenfors, *Knowledge in flux: Modeling the Dynamics of Epistemic States*, MIT Press, 1988.
[3] S. O. Hansson, 'A survey of non-prioritized belief revision', *Erkenntnis*, **50**, 413–427, (1999).
[4] J.-J. C. Meyer and W. van der Hoek, *Epistemic Logic for AI and Computer Science*, Cambridge University Press, 1995.
[5] L. Perrussel and J.M. Thévenin, 'A logical approach for describing (dis)belief change and message processing', in *Proceedings of AAMAS'04*, pp. 614–621. IEEE C.S., (2004).
[6] L. Perrussel, J.M. Thévenin, and T. Meyer, 'Mutual enrichment for agents through nested belief change: A semantic approach', in *Proc. of NMR'06*, (2006).