

Mutual Enrichment Through Nested Belief Change

Laurent Perrussel
IRIT–Université Toulouse 1
Manufacture des Tabacs, 21
allée de Brienne, F-31042
Toulouse Cedex - France
laurent.perrussel@irit.fr

Jean-Marc Thévenin
IRIT–Université Toulouse 1
Manufacture des Tabacs, 21
allée de Brienne, F-31042
Toulouse Cedex - France
thevenin@univ-tlse1.fr

Thomas Meyer^{*}
NICTA and UNSW
223 Anzac Parade
Kensington, NSW, 2052,
Australia
thomas.meyer@nicta.com.au

ABSTRACT

We investigate 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 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 the other agent's beliefs, as well as the preferences of the other agent. The main objective of nested beliefs is to improve cooperation between agents. We propose a logical framework for the acquisition process of nested beliefs and preferences. This acquisition process is the first step toward the elaboration of sophisticated interaction protocols. In this short paper we provide an informal outline of the framework, guided by an intuitive running example.

Keywords

multi-agent systems, reasoning about action, belief revision

1. INTRODUCTION

In a multi-agent context *nested beliefs* represent what agents believe about the beliefs of other agents. Nested beliefs improve interaction and cooperation among agents [5, 2, 10, 14]. Few papers deal with the problem of how agents acquire nested beliefs [9, 4]. None of them considers this in the context of cooperative dialogues. We propose an acquisition process for nested beliefs which describes how agents construct nested beliefs as they receive information. This provides the foundation on which we can build sophisticated dialogues or cooperation protocols.

^{*}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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.
Copyright 2006 ACM 1-59593-303-4/06/0005 ...\$5.00.

We consider multi-agent dialogues based on the *tell* KQML performative [6] which enables an agent to send a message to inform a receiver agent that a statement holds in its own belief base. We suppose that the sender provides the *source* of this statement [12]. The receiver may accept or ignore the incoming statement leading to non-prioritised revision of basic beliefs [8]. To perform revision agents need to be able to define preferences over their basic beliefs and thus the incoming statement may be considered as less preferred than conflicting basic beliefs. Preferences over basic beliefs is obtained by maintaining preferences over agents at the origin of basic beliefs, representing the reliability of the statement origin.

Based on this we define the acquisition of nested beliefs for the receiver of a *tell* performative. Regardless of the acceptance of the *tell* performative, the receiver is informed that the sender believes the incoming statement. After a *tell* performative the receiver performs non-prioritised revision on its nested beliefs about the sender. Since the sender provides the source of its beliefs, the receiver can deduce the nested preferences of the sender. The sender can also enrich its nested beliefs in several ways. For this, we introduce two performatives *accept* and *deny* which enable the receiver agent to inform the sender whether or not it actually changed its own beliefs after a *tell*. Using these performatives the sender can enrich its nested beliefs in several ways. After an *accept* the sender believes that the receiver believes in the content of the *tell*. If, during the revision process, an inconsistency occurs in the nested beliefs of the receiver, the sender can refine its knowledge about the preferences of the receiver. This is also the case after a *deny* statement.

To illustrate the issues involved, consider the following example. Agent Peter receiving messages from agent Paul and the police department. Suppose Peter considers Paul to be less reliable than the police department. At time 0, Paul tells Peter that *John is a murderer* (origin Paul). At time 1 Peter accepts the message. At time 2, Paul tells Peter that *if John is a murderer, John will go to jail* (origin Paul). Peter accepts the message and consequently Peter believes that *John will go to jail* at time 3. In addition, Peter believes that Paul believes both statements. After the second accept, Paul also believes that Peter believes both statements. At time 4, the police department tells Peter that *John is not a murderer* (origin Police). Because of its preferences, Peter accepts message at time 5 and now believes *John is not a murderer*. In addition, Peter believes that the police department believes *John is not a murderer*. After the accept message, the police believes that Peter believes *John is not a murderer*. Because it believes that Paul still believes that *John is a murderer*, at time 6, Peter tells Paul that *John is not a murderer* and the origin of this information is the police. Whether or not Paul accepts Peter's message, Paul has to change its nested beliefs about Peter. Paul now believes that Peter believes *John is not a murderer* and that Peter considers the police to be more reliable than himself.

The paper is structured as follows. Section 2 sketches a logical framework for reasoning about nested beliefs, preferences and the performatives *tell*, *accept*, and *deny*. Sections 3 and 4 give an overview of the mutual enrichment for the receiver and the sender. We conclude with a discussion on related, and future work.

2. THE LOGICAL FRAMEWORK

The *belief state* of an agent contains basic beliefs as well as nested beliefs. The basic beliefs of an agent a , denoted by B_a , are represented as signed statements, each consisting of a propositional sentence and the agent id of the source of the statement. Every agent a also maintains a set of signed statements for each of the other agents, representing what each of them currently believes, according to a (the nested beliefs of a). $B_{a,b}$ denotes what agent b believes according to agent a . We also assume that agents have preferences, arranged as a total preorder, over the set of agents A which describe the reliability of the sources of information [3, 12]. So $b \preceq_a c$ means: agent c is at least as preferred as b for agent a ($b \prec_a c$ and $b =_a c$ have the obvious meanings). Agents can also handle nested preferences. $c \preceq_{a,b} d$ means: agent a believes that for agent b agent d is at least as preferred as c . Together, these constitute the preference state of the agent.

We first present the language $\mathcal{L}_1$ for reasoning about basic beliefs and preferences. $\mathcal{L}_1$ is a restricted first order language which enables quantification over agent ids. It has a modal-like operator Bel for basic beliefs. $Bel(a, \phi_0, t)$ means agent a believes ϕ_0 signed by t . We limit agent beliefs to propositional sentences. To represent preferences we introduce the notation $b \leq_a c$ which stands for: agent a considers agent c to be at least as preferred as b . We extend $\mathcal{L}_1$ to the language $\mathcal{L}_1^K$ to deal with nested beliefs. Nested beliefs are restricted to $\mathcal{L}_1$ -formulae. We introduce the operator B to describe nested beliefs. Intuitively, $B(a, Bel(b, \phi_0, s))$ means agent a believes that agent b believes ϕ_0 signed by s , while $B(a, c \leq_b d)$ means that agent a believes that agent b considers agent d to be at least as preferred as c . So, in our example, at time 3 agent *peter* has the following belief state:

$$\begin{aligned} B_{peter} &= \{\langle murd, paul \rangle, \langle murd \rightarrow jail, paul \rangle\} \\ B_{peter, paul} &= \{\langle murd, paul \rangle, \langle murd \rightarrow jail, paul \rangle\} \\ B_{peter, police} &= \emptyset \end{aligned}$$

Considering that *peter* does not know *paul* and *police*'s preferences at the beginning, the preference state of *peter* is:

$$\begin{aligned} paul &\prec_{peter} peter \prec_{peter} police \\ paul &=_{peter, paul} peter =_{peter, paul} police \\ paul &=_{peter, police} peter =_{peter, police} police \end{aligned}$$

We now formalise the communication between agents. We assume that agents exchange information about their basic beliefs. We represent interactions between agents using three performatives. $Tell(s, r, \phi_0, t)$ stands for: agent s informs r that it believes ϕ_0 signed by t according to the standard KQML semantics [6]. When receiving a *Tell* performative, agent r performs non-prioritised revision according to its preferences [8]. $Accept(r, s, p)$ stands for: agent r informs s that it accepts the *Tell* performative p . Let $p = Tell(s, r, \phi_0, t)$. Then $Accept(r, s, p)$ means that agent r has revised its basic beliefs by $\langle \phi_0, t \rangle$ and thus believes ϕ_0 . $Deny(r, s, p)$ stands for: agent r informs s that it refuses to process the *Tell* performative p . If performative $p = Tell(s, r, \phi_0, t)$ it means that agent r has not revised its basic beliefs by $\langle \phi_0, t \rangle$.

The dynamics of the system is given by performatives. At each step one agent receives a performative and may change its state

accordingly, i.e. performatives are viewed as transition functions. We extend $\mathcal{L}_1^K$ to a language $\mathcal{L}_2$ to handle performative actions by considering a modal-like action operator $[p]\phi$, which states that after performative action p , ϕ is true. With $\mathcal{L}_2$ we can specify the behaviour of agents when they interact. For instance, we can express the honesty postulate as follows.

$$[Tell(s, r, \phi_0, t)]\top \rightarrow Bel(s, \phi_0, t) \quad (\text{Hon})$$

This postulate states that if a *Tell* performative occurs, then at the same time the sender believes the corresponding signed statement. Postulates such as (Hon) place constraints on the logical framework, ensuring that it behaves as required. Due to lack of space we shall, in the rest of the paper, only provide informal descriptions of the postulates required for our framework. The first such postulate, (KeNB1), requires that the nested beliefs and preferences of agents not involve in a dialogue do not change, and that the dialogue between the sender and the receiver is a two step process: first a *Tell* and then an *Accept* or a *Deny*.

3. DYNAMICS OF THE RECEIVER

In this section, we discuss the dynamics of the nested beliefs and nested preferences of agent r for the $Tell(s, r, \phi_0, t)$ performative, and then consider the acknowledgement sent by agent r to agent s . The nested beliefs and preferences of all agents except s have to remain unchanged. (KeNB2) propagates them to the next state and ensures that no new nested beliefs or preferences appear. Also, agent r changes its beliefs about s , whether or not it ignores the input statement. Because agents are supposed to be honest, agent r performs a standard prioritised revision of its nested beliefs about s by signed statement $\langle \phi_0, t \rangle$. (AdNB1) states that after the *Tell* performative r believes that s believes ϕ_0 and thus does not believe $\neg\phi_0$ anymore (because belief states are consistent). (ProNB1) states that all nested beliefs which hold after the *Tell* performative also hold at the previous state or are consequences of the new nested belief $B(r, Bel(s, \phi_0, t))$. (AdNB1) and (ProNB1) describe the expected behaviour of a revision operator for the nested beliefs but do not specify which nested beliefs have to be removed to facilitate the removal of $\neg\phi_0$ from the nested beliefs of r , since we do not insist on a specific method for revising nested beliefs.

(KeNP1) states that nested preferences about all agents except s are unchanged: it propagates them to the next state and ensures that no new preference appears. If agent r currently believes that s believes $\neg\phi_0$ signed by t' then it refines the s preferences as follows. According to the honesty principle and (AT) (see below), t is better than t' . (AdNBP1) formalises this intuition. (KeNP2) requires that the nested preferences of agent s about all agents except t are unchanged. But if r does not believe that s believes $\neg\phi_0$ then it cannot draw conclusion about agent s preferences. Agent r can refine preferences only if it faces inconsistencies during the revision process. (KeNP3) states that if r does not believe that s believes $\neg\phi_0$ then all nested preferences about s are unchanged.

This brings us to the receiver's acknowledgement. Whenever agent r receives $Tell(s, r, \phi_0, t)$, r determines if its beliefs are consistent with ϕ_0 , and if not, which beliefs have to be dropped by taking into account its preference relation. Due to the non-prioritised aspect the receiver may deny or accept the *Tell* performative. If agent r believes $\neg\phi_0$ and the signatures of $\neg\phi_0$ are at least as preferred as t (the signature of ϕ_0), postulate (DT) requires that agent r deny the *Tell* performative. Otherwise there is no conflict or t is strictly more preferred than all the signatures of $\neg\phi_0$, and postulate (AT) requires that r accept the *Tell* performative.

4. DYNAMICS OF THE SENDER

In this section we describe the dynamics of the sender's nested beliefs and preferences with respect to the performatives *Accept* and *Deny*. (KeNB3) states that the nested beliefs of agents not involved in the *Accept* or *Deny* performative do not change. Next, we focus on the nested beliefs about r . According to (AT), if r accepts the message of s , agent s believes that r believes ϕ_0 and thus does not believe $\neg\phi_0$, as specified by (AdNB3). According to (DT), if r refuses the *Tell* performative then s concludes that r believes $\neg\phi_0$, as specified by (AdNB2). Like (ProNB1), postulates (ProNB2) and (ProNB3) state that all nested beliefs holding at the next state also hold at the current state, or are consequences of the new nested belief $B(r, \text{Bel}(s, \phi_0, t))$. So in both cases, s nested beliefs about r undergoes a prioritised revision: by $\langle\phi_0, t\rangle$ in the case of *Accept*, and by $\langle\neg\phi_0, t\rangle$ in the case of *Deny*.

This brings us to the sender's nested preferences. (KeNP4) states that the sender's nested preferences about agents other than r do not change. Next, agent s draws conclusions about t depending on whether r accepts or refuses a *Tell* performative. Whenever agent r accepts, agent s refines its nested preferences only if s currently believes that r believes $\neg\phi_0$. In that context s refines agent r 's preferences. (AdNP2) requires that t be strictly more preferred than the signatures of $\neg\phi_0$ if r accepts the performative *Tell*(s, r, ϕ_0, t). (AdNP3) requires that if r refuses 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 t . Finally, (KeNP5) and (KeNP6) deal with those nested preferences, about r , of s that do not change. They are similar to (KeNP2) and (KeNP3).

We now conclude our example. At time 4, the *police* tells *peter* that *John* is not a murderer, and *peter* accepts it.

$$\begin{aligned} p_4 &= \text{Tell}(\text{police}, \text{peter}, \neg\text{murd}, \text{police}) \\ p_5 &= \text{Accept}(\text{peter}, \text{police}, p_4) \end{aligned}$$

By (AdNB1) and (AdNB3) we get the new nested beliefs:

$$\begin{aligned} [p_4]B(\text{peter}, \text{Bel}(\text{police}, \neg\text{murd}, \text{police})) \\ [p_5]B(\text{police}, \text{Bel}(\text{peter}, \neg\text{murd}, \text{police})) \end{aligned}$$

At time 1 *peter* accepted *paul*'s message that *John* is a murderer (*murd*) which gives

$$B(\text{paul}, \text{Bel}(\text{peter}, \text{murd}, \text{paul}) \wedge \text{paul} =_{\text{peter}} \text{police})$$

according to (KeNB1). At time 6, this nested belief is still true according to the accepted behaviour of nested beliefs given by (KeNB1), (AdNB3), (ProNB2) and (KeNP6). At time 6 *peter* tells *paul* that *John* is not a murderer: $p_6 = \text{Tell}(\text{peter}, \text{paul}, \neg\text{murd}, \text{police})$.

According to (AdNB1) and (AdNP2) we obtain the following new nested beliefs and preferences for *paul*:

$$\begin{aligned} [p_6]B(\text{paul}, \text{Bel}(\text{peter}, \neg\text{murd}, \text{police})) \\ [p_6]B(\text{paul}, \text{paul} <_{\text{peter}} \text{police}) \end{aligned}$$

So, at stage 6 *paul* has refined its nested preferences about *peter*. At the beginning of the dialogue *paul* does not know whether *peter* prefers *paul* or the *police*. At stage 6, *paul* believes that *peter* prefers the police to himself.

5. CONCLUSION

We have outlined a framework for handling the dynamics of nested beliefs and preferences in the context of agent interactions. We proposed a logical framework which allows for the expression of nested beliefs, preferences, and performatives, and showed how agents acquire nested beliefs and preferences. For this we

have specified the expected behaviour of agents handling the *Tell*, *Accept* and *Deny* performatives.

Related work includes that of [4], which proposes a logic-based definition for nested belief change and mental state construction related to agent-communication languages. Our proposal differs from theirs in that it deals with preferences of agents. Our work is also closely related to [13, 12] where it is shown how to embed basic belief change in a logical system. We extend this by considering nested beliefs and preferences. Finally, our work is loosely related to the dynamics of preferences [1]. Our contribution is to formulate this dynamics in terms of agent interactions. We plan to extend our framework by considering nested *disbeliefs* [7]. Explicit disbeliefs allow for a more nuanced type of interaction between agents and provide new opportunities for the discovery of nested beliefs and preferences. We also plan to revisit persuasion protocols [11] so that agents can generate smarter arguments.

6. REFERENCES

- [1] G. Brewka. Well-Founded Semantics for Extended Logics Programs with Dynamic Preferences. *Journal of Artificial Intelligence Research*, (4):19–36, 1996.
- [2] P. Cohen and H. Levesque. Rational Interaction as the Basis for Communication. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*, pages 221–256. MIT Press, 1990.
- [3] A. Dragoni and P. Giorgini. Revising beliefs received from multiple sources. In M. Williams and H. Rott, editors, *Frontiers of Belief Revision, Applied Logic*. Kluwer, 1999.
- [4] A. Dragoni, P. Giorgini, and L. Serafini. Mental States Recognition from Communications. *J. Logic Computat.*, 12(1):119–136, 2002.
- [5] A. C. F. Grasso and R. Jones. Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition. *International Journal of Human-Computer Studies*, 53(1077–1115), 2000.
- [6] T. Finin, Y. Labrou, and J. Mayfield. KQML as an agent communication language. In J. Bradshaw, editor, *Software Agents*. MIT Press, 1997.
- [7] A. Ghose and R. Goebel. Belief states as default theories: studies in non-prioritized belief change. In H. Prade, editor, *ECAI'98*, pages 8–12, 1998.
- [8] S. O. Hansson. A survey of non-prioritized belief revision. *Erkenntnis*, 50:413–427, 1999.
- [9] A. Herzig and D. Longin. Belief dynamics in cooperative dialogues. *J. of Semantics*, 17(2), 2000. vol. published in 2001.
- [10] D. Kinny, M. Ljungberg, A. S. Rao, E. Sonenberg, G. Tidhar, and E. Werner. Planned team activity. In C. Castelfranchi and E. Werner, editors, *Artificial Social Systems – MAAMAW'92*, volume 830 of *LNAI*, pages 226–256. Springer-Verlag, 1992.
- [11] S. Parsons, M. Wooldridge, and L. Amgoud. An analysis of formal inter-agent dialogues. In M. Gini, T. Ishida, C. Castelfranchi, and W. L. Johnson, editors, *Proceedings of AAMAS'02*, pages 394–401. ACM Press, July 2002.
- [12] L. Perrussel and J. Thévenin. A logical approach for describing (dis)belief change and message processing. In *AAMAS'04*, pages 614–621. IEEE C.S., 2004.
- [13] J. Roorda, W. van der Hoek, and J. Meyer. Iterated Belief Change in MAS. In C. Castelfranchi and W. Johnson, editors, *AAMAS'02*, pages 889–896. ACM Press, July 2002.
- [14] D. Sperber and D. Wilson. *Relevance*. Harvard University Press, 1986.