Generalised Proof-Theory for Multi-Agent Autoepistemic Reasoning

Yongyuth Permpoontanalarp
Department of Computing, Imperial College
London SW7 2BZ, England
ja8@doc.ic.ac.uk

John Yuejun Jiang
Pacific Telesis Technology Laboratory
5000 Executive Parkway, 333
San Ramon, CA 94583
jjiang@ttl.pactel.com

Abstract

Over the past few years, several different approaches have been proposed to deal with multi-agent autoepistemic reasoning. Despite some limitations, each approach is important in its own right. While Parikh's approach allows an agent to reason nonmonotonically about other agents' knowledge, it cannot reason about other agents' nonmonotonic reasoning. Morgenstern's logic provides a limited way to deal with the problem but unfortunately it is not constructive. Although Halpern introduces an algorithmic definition of multi-agent nonmonotonic reasoning, his approach cannot deal with default reasoning even for the single-agent case. The purpose of this paper is to propose an integrated theory that deals with these problems. Using examples from speech acts theory, we demonstrate some unintuitive results against existing approaches. We then develop a simple and yet generalised proof-theoretic framework with constructive interpretation for multi-agent autoepistemic reasoning. We show that this framework retains the advantages of existing approaches but does not have their peculiar results. Surprisingly, the new proof-theoretic framework can be obtained by a simple modification of Parikh's approach. Furthermore, the results show that our framework generalises Morgenstern's approach.

Introduction

Autoepistemic (AE) logic [Moore 1985] is a nonmonotonic logic about an implicit rational agent who reasons about his beliefs and ignorance by introspection. Nonmonotonicity is accommodated in AE logic by the property that a conclusion based on what is currently believed or disbelieved can be withdrawn in the presence of new beliefs acquired by the implicit agent. Despite its association with beliefs, it is however quite difficult to extend AE logic to reason about multi-agent beliefs. It is felt that multi-agent autoepistemic logic can be useful to many application domains such as cooperative planning, electronic commerce, speech acts [Perrault 1987, Appelt and Konolige 1988]. A clear understanding and a constructive proof mechanism for such a logic is therefore worth of investigation.

With the explosion of world-wide-web, electronic commerce is also coming into existence. While the initial concern is perhaps on security (Secure HTTP or SSL), there will be a time soon when transactions need be carried out by representative agents of users that deal with intention recognition and negotiation. For example, if John's agent tries to get the cheapest airfare from some travel agency, a dumb agent of the agency can simply offer the cheapest airfare. But a clever agent equipped with commonsense reasoning can offer John a great deal which may not the cheapest (eg. flying British Airways which is only 50 dollar more expensive than the cheapest airfare by Russian Airline). Such transactions will involve multi-agent nonmonotonic reasoning.

Speech acts are concerned with the effects of utterances on the mental states of speakers and hearers. Recently, a paradigm of Software Agents based on speech acts [Genesereth and Ketchpel 1994] has emerged. Its aim is to support the seamless interoperation amongst heterogeneous agents. A major key to achieve the aim is a common Agent-Communication language (ACL) [Finin et al. 1994, Genesereth et al. 1992]. However, current proposal [Labrou and Finin 1994] on the semantics of Knowledge Manipulation and Query Language (KQML), which is a major part of ACL, does not provide a formal treatment on incomplete beliefs amongst agents. Because ignorance is often based on assumptions, a logic on multi-agent nonmonotonic reasoning about ignorance such as a multi-agent autoepistemic logic would be a particularly suitable formalism for characterising agent communication languages.

Over the past few years, several different approaches [Konolige 1988a, Morgenstern 1990, Parikh 1991, Lakemeyer 1993, Halpern 1993, Jiang 1994] have been proposed to deal with multi-agent autoepistemic reasoning. Despite some limitations, each approach is important in its own right.

Konolige [Konolige 1988a] proposed Hierarchic AE (HAE) logic to deal with attitude revision in speech acts. This logic is a restricted but constructive variant of Moore's AE logic [Moore 1985]. It consists of a hierarchic structure of theories in which introspection at a level is only defined with respect to theories at a lower level. It supports priority reasoning by explicit level distinctions. However, HAE can be quite cumbersome on reasoning about dynamic beliefs since adding any new information involves choosing a correct level of theory.

Later on, Parikh [Parikh 1991] introduced LK logic. This logic does not have the levels of HAE and yet is still constructive. However, LK logic seems to yield some unintuitive results for default reasoning. In addition, it only allows an agent to reason nonmonotonically about other agents' knowledge but not about other agents' nonmonotonic reasoning. Morgenstern [Morgenstern 1990] provides a limited way to deal with the problem but unfortunately it is not constructive. Furthermore, Morgenstern's Multi-Agent NonMonotonic Logic (MANML) seems to suffer from peculiar results in which Moore's AE logic [Moore 1985] does not have. More precisely, MANML can have two extensions for reasoning in taxonomic hierarchies while its single-agent counterpart, Moore's AE logic, has only one extension.

Although Halpern [Halpern 1993] developed an algorithmic definition of multi-agent nonmonotonic reasoning, his approach cannot deal with default reasoning even for the single-agent case. Recently, Lakemeyer [Lakemeyer 1993] extended Levesque's logic of all I know [Levesque 1990] to handle multi-agent case and the logic is called "all they know". The logic of all they know shares some similar notions to Halpern's approach [Halpern 1993]. However, it does not have a complete proof theory.

The purpose of this paper is to propose an integrated multi-agent autoepistemic logic that deals with these problems. Although the logic is not directly applied to speech acts theory or agent communication languages, we will use examples from speech acts theory to motivate our works and to demonstrate some unintuitive results against existing approaches. Unlike many works on logics which are usually semantics-based, our work is proof-theoretic in the sense that we only give an algorithmic definition of our logic. We consider this as an important step to study applications of multi-agent nonmonotonic reasoning.

The proposed proof theory is a generalisation of Parikh's approach [Parikh 1991] with a more generic rule of inference for nonmonotonic reasoning and a more conservative modal system (S4F). We show that our theory can solve many of the peculiar problems in existing approaches. We also show that our framework generalises Morgenstern's works [Morgenstern 1990] on principles of arrogance.

This paper is organised as follows. We begin with some preliminary background followed by some issues about representation for multi-agent reasoning in brief. Then, we give an account on speech acts theory with some motivating examples for analysing existing approaches. As a demonstration of the usefulness of Autoepistemic reasoning in software agents, we show its relation to Agent-Communication language. Following a detail discussion on previous approaches and their problems, a new generalised proof theory is shown. Finally, some results on the new proof theory are given.

Preliminary background

Definition 1 *$S5_n$ modal system is a normal modal system which consists of the following axioms and rules of inference.*

1. *All tautologies.*
2. *K axiom:* $\mathbf{K}_i(\alpha) \land \mathbf{K}_i(\alpha \rightarrow \beta) \rightarrow \mathbf{K}_i\beta$
3. *T axiom:* $\mathbf{K}_i\alpha \rightarrow \alpha$
4. *4 axiom:* $\mathbf{K}_i\alpha \rightarrow \mathbf{K}_i\mathbf{K}_i\alpha$
5. *5 axiom:* $\neg\mathbf{K}_i\alpha \rightarrow \mathbf{K}_i\neg\mathbf{K}_i\alpha$
6. *Modus ponens:* $\frac{\alpha \quad \alpha\rightarrow\beta}{\beta}$
7. *Necessitation rule of inference:* $\frac{\vdash\phi}{\vdash\mathbf{K}_i\phi}$

Definition 2 *$S4F_n$ modal system is a normal modal system which consists of the following axioms and rules of inference:*

1. *All tautologies.*
2. *K,T,4 axioms*
3. *F axiom:* $\neg\mathbf{K}_i\neg\alpha \land \neg\mathbf{K}_i\neg\mathbf{K}_i\beta \rightarrow \mathbf{K}_i(\neg\mathbf{K}_i\neg\alpha \lor \beta)$
4. *modus ponen and necessitation rule.*

Definition 3 *$SW5_n$ modal system is a normal modal system which consists of the following axioms and rules of inference:*

1. *All tautologies.*
2. *K,T,4 axioms*
3. *W5 axiom:* $\alpha \rightarrow (\neg\mathbf{K}_i\neg\mathbf{K}_i\alpha \rightarrow \mathbf{K}_i\alpha)$
4. *modus ponen and necessitation rule.*

Other modal system can also be obtained similarly. For instance, KD45 system consists of all tautologies, K,4,5 axioms together with D axiom which is $\mathbf{K}_i\alpha \rightarrow \neg\mathbf{K}_i\neg\alpha$, modus ponen and necessitation rule.

Definition 4 *A Generalised Multi-Agent Autoepistemic (GMAE) language is a language of the Propositional logic augmented by modal operators $\mathbf{K}_i$ for each of agents i where $1 \leq i \leq n$. The GMAE is defined as inductively as follows.*

1. *a propositional atom is a GMAE formula.*
2. *if ϕ, ψ are GMAE formulae, so are $(\phi \lor \psi)$, $(\phi \land \psi)$, $\neg\phi$ and $\neg\psi$.*
3. *if ϕ is a GMAE formula, so is $\mathbf{K}_i\phi$ for each agent i.*

Objective formulae are formulae in which there is no occurrence of any modal operator while knowledge formulae are formulae of the form $\mathbf{K}_i\phi$ where ϕ is any kind of formulae. The dual of $\mathbf{K}_i\phi$ is $\neg\mathbf{M}_i\neg\phi$. While $\mathbf{M}_i\phi$ should be read as "it is possible for agent i to know ϕ", $\mathbf{K}_i\phi$ should be read as "agent i knows ϕ".

Representation forms for Multi-Agent reasoning

There are two forms of representation for multi-agent reasoning [Jiang 94]. The *introspective* form associates each agent with a theory that defines its beliefs about itself and about other agents' beliefs. For example, such a theory for agent A could be $\{\mathbf{K}_B p \vee \mathbf{K}_C q, r\}$. The *extrospective* form is simply a theory consisting of a God (external observer)'s view about all agents' beliefs about themselves. For example, such a theory could be $\{\mathbf{K}_A(\mathbf{K}_B p \vee \mathbf{K}_C q), \mathbf{K}_A r, \mathbf{K}_B s\}$.

For single-agent case, both representations do not have much difference. Nevertheless, this is not the case for multi-agent case. The first allows only one agent to express its own reasoning while the second provides a God (external observer)'s view of nonmonotonic reasoning amongst agents.

Speech Acts theory

Speech acts theory regards utterances as actions that transform the mental states of speakers and hearers. It was argued [Appelt and Konolige 1988] that the reasoning involved often deals with multi-agents' beliefs and ignorance about themselves and each other. Due to its natural way of reasoning about agents' ignorance, multi-agent autoepistemic logic therefore can be seen to play important roles in speech acts in many aspects.

Perrault [Perrault 1987] applied Reiter's default logic [Reiter 1980] to reason about speech acts. Subsequently, Appelt and Konolige [Appelt and Konolige 1988] proposed Hierarchic Autoepistemic (HAE) logic to reason about attitude revision in speech acts. However, the discussion about these frameworks is out of scope of this paper. Here we shall only consider an adaptation from an example in [Appelt and Konolige 1988] so that it can be used to demonstrate some peculiar results of previous approaches to multi-agent autoepistemic reasoning. Some general non-logical axioms for characterising the *tell* speech act are given below.

(1) $\text{tell}(s,h,\phi) \rightarrow \mathbf{W}_s\mathbf{K}_h\mathbf{K}_s\phi$
(2) $\mathbf{K}_i\mathbf{W}_i\phi \leftrightarrow \mathbf{W}_i\phi$
(3) $\text{tell}(s,h,\phi) \rightarrow \mathbf{K}_h(\neg\mathbf{K}_h\neg\mathbf{W}_s\mathbf{K}_h\phi \rightarrow \mathbf{W}_s\mathbf{K}_h\phi)$
(4) $\text{tell}(s,h,\phi) \rightarrow \mathbf{K}_h(\mathbf{W}_s\mathbf{K}_h\phi \wedge \neg\mathbf{K}_h\neg\phi \wedge \neg\mathbf{K}_h\neg\mathbf{K}_s\phi \rightarrow \phi)$

s,h denote speaker and hearer, respectively and $\mathbf{W}_i p$ means that an agent i wants p. While the first axiom deals with the speaker's mental state for making an utterance, the second axiom makes the inter-connection between knowledge and desire. The third axiom says that after being told, a hearer becomes know that the speaker wants him to know about what the speaker told if it is consistent with the hearer's knowledge about the speaker's desire about the hearer's knowledge. The fourth axiom expresses that after being told, a hearer becomes know what the speaker told if he knows that the speaker wants him to know the utterance and the utterance is not conflict with the hearer's knowledge and with what the hearer knows about the speaker's knowledge. Note that the last condition $(\neg\mathbf{K}_h\neg\mathbf{K}_s\phi)$ in the last axiom is for the hearer to detect if the speaker lies.

Speech Act theory vs. ACL

Recently, a paradigm of Software Agents based on speech acts [Genesereth and Ketchpel 1994] has emerged. Its aim is to support the seamless interoperation amongst heterogeneous agents. A major key to achieve the aim is a common Agent-Communication language (ACL) [Finin et al. 1994, Genesereth et al. 1992]. However, current proposal [Labrou and Finin 1994] on the semantics of Knowledge Manipulation and Query Language (KQML), which is a major part of ACL, does not provide a formal treatment on incomplete beliefs amongst agents. Because ignorance is often based on assumptions, a logic on multi-agent nonmonotonic reasoning about ignorance such as a multi-agent autoepistemic logic would be a particularly suitable formalism for characterising agent communication languages. Here we show some examples of the current proposal [Labrou and Finin 1994] which reveals the need of a formal reasoning based on agents' ignorance and incomplete information.

- The completion condition of the speaker (agent A) to perform *tell(A,B,X)* performatives successfully is that agent B becomes knowing that agent A believe X. Such condition holds *unless* the speaker receives a message suggesting agent B's inability to acknowledge properly the *tell* performative. The word *unless* suggests reasoning, performed by agent A, with incomplete information regarding the message sent by agent B to agent A to acknowledge the agent's B inability.
- The precondition of the speaker (agent A) to perform *ask-if(A,B,X) and ask-all(A,B,X)* performatives successfully is that the speaker agent A does not know about agent's B belief regarding X. Apparently, this involves a reasoning performed by agent A about agent A's ignorance about agent B's belief.

Previous approaches

Morgenstern's MANML

The language of multi-agent NonMonotonic Logic (MANML) [Morgenstern 1990] is just like the Generalised Multi-Agent Autoepistemic (GMAE) language except that it uses a modal operator $\mathbf{L}_i$. However, for simplicity reason, we shall use $\mathbf{K}_i$ instead.

In order for agents to perform Autoepistemic reasoning, Morgenstern gave a definition of MANML stable set as follows.

Definition 5 *A MANML stable set rules for an MANML theory T are defined as follows.*

1. *if $p_1, .., p_n \in T$ and $p_1, .., p_n \vdash q$, then $q \in T$. Here $\vdash$ denotes standard FOL derivability.*
2. *if $\mathbf{K}_i p_1, .., \mathbf{K}_i p_n \in T$ and $p_1, .., p_n \vdash q$, then $\mathbf{K}_i q \in T$.*
3. *if $\mathbf{K}_i p \in T$, then $\mathbf{K}_i \mathbf{K}_i p \in T$.*
4. *if $\mathbf{K}_i p \notin T$, then $\mathbf{K}_i \neg \mathbf{K}_i p \in T$.*

To enable multi-agent nonmonotonic reasoning, the Principle of Moderate Arrogance is introduced and it is meant to assume that agents are arrogant with respect to their beliefs about other agents' beliefs.

Definition 6 *A principle of Moderate Arrogance for an agent x toward agent y for an MANML theory T is defined as follows.*

If $\mathbf{K}_x\mathbf{K}_y(\mathbf{K}_y\alpha \wedge \neg\mathbf{K}_y\beta \rightarrow \gamma) \in T$, $\mathbf{K}_x\mathbf{K}_y\alpha \in T$ and $\mathbf{K}_x\mathbf{K}_y\beta \notin T$, then $\mathbf{K}_x\mathbf{K}_y\gamma \in T$.

Morgenstern also proposed another principle to capture an agent's circumspect reasoning about ascribing the absence beliefs to other agents.

Definition 7 *A principle of Cautious Arrogance for an agent x toward agent y for an MANML theory T is defined as follows.*

If $\mathbf{K}_x\mathbf{K}_y(\mathbf{K}_y\alpha \wedge \neg\mathbf{K}_y\beta \rightarrow \gamma) \in T$, $\mathbf{K}_x\mathbf{K}_y\alpha \in T$, $\mathbf{K}_x\mathbf{K}_y\beta \notin T$ and $\mathbf{K}_x\beta \notin T$, then $\mathbf{K}_x\mathbf{K}_y\gamma \in T$.

Note that Morgenstern considered the MANML stable set rules together with either the Principle of Moderate Arrogance or the Principle of Cautious Arrogance.

Example 1 *Let T1={$\mathbf{K}_i\mathbf{K}_j(true)$}. Since true $\vdash \mathbf{K}_j p \vee \neg\mathbf{K}_j p$, we have $\mathbf{K}_i\mathbf{K}_j(\neg\mathbf{K}_j p \rightarrow \neg\mathbf{K}_j p) \in T1$. Thus, $\mathbf{K}_i\mathbf{K}_j\neg\mathbf{K}_j p$ can be derived from the theory. Notice that $\mathbf{K}_i\neg\mathbf{K}_j p$ cannot be derived from the theory. This is rather peculiar since agent i believes that agent j believes that j does not believe p and yet it is not the case that agent i believes that j does not believe p.*

Problems of Morgerstern's MANML

1. The Principles of Arrogances given by Morgenstern for MANML are restricted to two levels of agents only. Even for two agents case, the theory is not generalised enough to allow an agent a to emulate b's reasoning about agent a itself, eg. $\mathbf{K}_a\mathbf{K}_b\mathbf{K}_a(\neg\mathbf{K}_a\phi \rightarrow \psi)$. Furthermore, the definitions stick to a certain interpretation of default reasoning [Konolige 1988b]. As a result, other natural interpretation of default reasoning, for example [Truszczynski 1991{a,b}], cannot be reasoned about in MANML.
2. MANML seems to produce contradictory results for some case in multi-agent reasoning. For instance, in speech acts, there is a situation called sincere assertion where a speaker utters sincerely his knowledge ($\mathbf{K}_s p$) and the utterance is consistent with the hearer's knowledge. Hence, the hearer should believe the speaker. Consider a theory T.

   $\mathbf{K}_s p \wedge$
   $tell(s, h, \mathbf{K}_s p) \wedge$
   $\mathbf{K}_h(\neg\mathbf{K}_h\neg\mathbf{K}_s p \wedge \neg\mathbf{K}_h\neg\mathbf{K}_s\mathbf{K}_s p \rightarrow \mathbf{K}_s p)$

   Indeed, $\mathbf{K}_h\mathbf{K}_s$(true) is inherently derivable in the theory. According to MANML stable set rule and the fact that $\vdash$ true $\rightarrow \neg\mathbf{K}_s p \vee \mathbf{K}_s p$, we thus derive $\mathbf{K}_h\mathbf{K}_s(\neg\mathbf{K}_s p \rightarrow \neg\mathbf{K}_s p)$. Therefore $\mathbf{K}_h\mathbf{K}_s\neg\mathbf{K}_s p$ is obtained as a result. However $\mathbf{K}_h\mathbf{K}_h\mathbf{K}_s p$ can also be derived from the third conjunct of the given theory. These two formulae are obviously contradictory to each other.
3. It seems that when collasping MANML to single-agent case, the framework may yield undesirable results for which Moore's single-agent AE logic [Moore 1985] does not have. What this means is that some desirable properties in single-agent autoepistemic logic may be lost in its multi-agent counterpart. Consider an example in [Gelfond 1988] about reasoning in taxonomic hierarchies.

   $\mathbf{K}_i$ thing(fred)
   $\mathbf{K}_i$(bird(X) $\rightarrow$ thing(X))
   $\mathbf{K}_i$ bird(tweety)
   $\mathbf{K}_i$(thing(X) $\wedge\neg\mathbf{K}_i$ ab(X,thing,fly) $\rightarrow$ ¬fly(X))
   $\mathbf{K}_i$(bird(X) $\wedge\neg\mathbf{K}_i$ ab(X,bird,fly) $\rightarrow$ fly(X))
   $\mathbf{K}_i$(bird(X) $\wedge\neg\mathbf{K}_i$ ab(X,bird,fly) $\rightarrow$ ab(X,thing,fly))

   The last formula is called "cancellation axiom" which is to ensure that specific information is preferred over those less specific. Unfortunately, two sets of results can be obtained in MANML which are {fly(tweety),¬fly(tweety)} and {fly(tweety)} but only the latter is equivalent to Moore's AE logic.
4. Although MANML allows the extrospective form of representation, it does not address nonmonotonic reasoning regarding disjunctions. For example, it is not clear what a theory like $\{\mathbf{K}_x p \vee \mathbf{K}_y q\}$ will lead to.

Parikh's LK Logic

The language of LK logic [Parikh 1991] is like the Generalised Multi-Agent Autoepistemic (GMAE) language and the LK logic is based on $\mathbf{S5}_n$ modal logic.

Definition 8 *LK-theory is a set of formulae which is closed under $\mathbf{S5}_n$ modal logic.*

Definition 9 *The rule* **M** *for nonmonotonic reasoning for a theory T is a rule of inference:*

$$\frac{T \nvdash \mathbf{K}_i\phi}{T \vdash \neg\mathbf{K}_i\phi}$$

The system NM consists of $\mathbf{S5}_n$ and the above **M** rule.

Definition 10 *A normal nonmonotonic proof of ϕ from a theory T in the system NM is a sequence of monotonic LK-theories T_k, $0 \leq k \leq m$ such that*

1. *T_0 is the LK-theory generated by T and*
2. *for each $k < m$, T_{k+1} is the LK-theory obtained by adding to T_k a formula $\neg\mathbf{K}_i\neg\alpha$ where*
   (a) *$\mathbf{K}_i\neg\alpha$ is not in T_k and*
   (b) *For all subformulae $\mathbf{K}_i\beta$ of α, either $\mathbf{K}_i\beta$ or $\neg\mathbf{K}_i\beta$ is already in T_k.*
3. *$\phi \in T_m$.*

Problems of Parikh's LK logic

1. Parikh's LK logic does not allow an agent to reason nonmonotonically about other agents' nonmonotonic reasoning. For example, for a theory T $\{K_iK_j(\neg K_j\neg fly \rightarrow fly)\}$ which expresses agent i's knowledge about agent j's default, LK logic cannot conclude $K_iK_j fly$. This is because LK logic cannot derive $K_i\neg K_j\neg fly$ but $\neg K_iK_j fly$ or $\neg K_i\neg K_j\neg fly$. Therefore, LK logic only allows an agent to reason nonmonotonically about other agents' knowledge.
2. LK logic seems to yield contradictory results for some cases in multi-agent reasoning. For example, in speech acts, there is a situation called insincere assertion where a speaker lies to the hearer and the hearer can detect the lie so the hearer should not believe the speaker as a result. Consider a theory T.
   $\neg K_h p \wedge$
   $tell(s, h, K_sK_hp) \wedge$
   $K_h(\neg K_h\neg K_sK_hp \wedge \neg K_h\neg K_sK_sK_hp \rightarrow K_sK_hp)$
   Since the hearer knows that he does not know about p, he should be able to detect when the speaker lies (K_sK_hp) that the hearer knows p. However, both $\neg K_h\neg K_sK_hp$ and $\neg K_h\neg K_sK_sK_hp$ can be proved. As a result, $K_hK_sK_hp$ is derived and it yields K_hK_hp which contradicts to $K_h\neg K_hp$.
3. LK logic also suffers the same problem as Morgenstern's MANML [Morgenstern 1990] in (3) that for the single-agent case, LK logic may not behave the same as its single-agent AE counterpart, Moore's AE logic [Moore 1985]. More precisely, in the example (3), LK logic can yield two sets of results which are the same as MANML.
4. Default reasoning even for the single-agent case can have two extensions. For instance, a theory $\{K_j(\neg K_jp \rightarrow q)\}$, which is a default about q in the absense of the information regarding p, has two extensions: K_jp and K_jq. The former is unintuitive. Also, in $\{K_j(\neg K_j\neg fly \rightarrow fly)\}$, which is a default about fly, either $K_j fly$ or $\neg K_j fly$ can be proved and the latter is obviously peculiar. Note that although other modal system such as **S4F** or **SW5**, is used instead of **S5**, these anomalies still remain.

Halpern's HM-ONL

The language of HM-ONL logic [Halpern 1993] is just like GMAE but it is based on **KD45** or **K45** logic. Halpern defined a notion of only knowing with many agents and he also gave a semantic and recursive syntactical characterisation for HM-ONL logic.

Definition 11 *A set $\mathcal{D}^i_S(\alpha)$ consists of all the formulae agent i knows given that agent i knows only α under S modal logic and the set is defined as follows.*

$$\varphi \in \mathcal{D}^i_S(\alpha) \text{ iff } \models_S (K_i\alpha \wedge \varphi^{\alpha,i}) \rightarrow K_i\varphi$$

where $\varphi^{\alpha,i}$ is the conjunction of $K_i\psi$ for all subformulae $K_i\psi$ of φ for which $\psi \in \mathcal{D}^i_S(\alpha)$, and $\neg K_i\psi$ for all subformulae $K_i\psi$ of φ for which $\psi \notin \mathcal{D}^i_S(\alpha)$ where φ is considered a subformula of itself.

Theorem 1 *For $S \in \{\mathbf{KD45}_n, \mathbf{K45}_n\}$, the formula α is i-honest iff $\mathcal{D}^i_S(\alpha)$ is propositionally consistent.*

Problems of HM-ONL

1. The definitions given by Halpern do not allow nested nonmonotonic reasoning of an agent about other agents' nonmonotonic reasoning, for instance, $\{K_iK_j(\neg K_j\neg fly \rightarrow fly)\}$. However, Halpern's ONL can handle nested knowledge formulae only for the single-agent case. For instance, a theory $\{K_j(\neg K_j\neg K_jp \rightarrow K_jp)\}$ does not have K_jp as a result.
2. HM-ONL cannot deal with default reasoning even for single-agent case. For example, a theory T1 $\{K_i(\neg K_ip \rightarrow q)\}$, which is a default about q in the absence of the information regarding p, derives K_iq and $\neg K_iq$. As a result, the default cannot be reasoned about.
3. ONL can lead to inconsistency for some cases. As noticed by Halpern, a theory T2 $\{K_ip \vee K_iq\}$ can yield $\neg K_ip \wedge \neg K_iq$ which contradicts to T2.

Although Halpern was interesting in the notion of i-honest for defining the algorithmic characterisation, Halpern's syntactical characterisation is not adequate to deal with multi-agent reasoning in general as we shown above.

Jiang's extrospective MAE logic

Jiang's multi-agent autoepistemic (MAE) logic [Jiang 1994] is a multi-agent counterpart of Moore's Autoepistemic (AE) logic which is for single-agent case. The language of MAE logic is based on two modal operators **L** and Bel. The intended meaning of **L** is the provability while Bel is to indicate the scope of beliefs for particular agent. However, for simplicity reason, we shall use the modal operator **K** instead of **L**.

Jiang proposed two logics: AE^{Epis} and MAE. Indeed, AE^{Epis} logic can be obtained by replacing the ordinary part of Moore's AE logic with some epistemic logic (Epis) which is a monotonic modal logic of beliefs. While AE^{Epis} logic allows an external observer or God to express the representation about other agents, MAE, on the other hand, allows agents to express the representation about other agents through the external observer's reasoning. However, Jiang showed that both logics are essentially equivalent by some transformation.

Definition 12 *The language of AE^{Epis} can be defined as follows.*

1. *a propositional atom is an AE_{Epis} formula.*
2. *if ϕ and ψ are AE_{Epis} formulae, so are $\neg\phi, \phi\vee\psi, \phi\wedge\psi$ and $\phi \rightarrow \psi$.*
3. *if ϕ is an AE_{Epis} formula, then $K\phi$ is AE_{Epis} formula.*

4. *if ϕ is an ordinary AE_{Epis} formula, i.e. it does not contain a* **K** *operator, then Bel(agent,ϕ) is an AE_{Epis} formula. Here agent is a member of the set of agents in AE_{Epis}.*

Definition 13 *Let T be an AE^{Epis} theory. Let $Cn_{(S,U)}$ denote the consequential closure of S and U where S is any monotonic modal logic on the modal operator* **K** *and U is any monotonic epistmic logic on the modal operator Bel. Then E is a (S, U)-extension of T iff*

$$E = Cn_{(S,U)}(T \cup \{\neg \mathbf{K}\phi \mid \phi \notin T)$$

Comparison

1. Jiang's extrospective MAE logic allows a MAE theory to be defined against nested agents' belief space but uses two modal operators. Thus, the framework allows the second form of multi-agent knowledge representation which is discussed in earlier section.
2. Jiang's extrospective MAE logic is based on Self-referential nonconstructive fixpoint definitions. It can perform default reasoning in multi-agent case properly. For example, $\{\mathbf{K}_i(\neg\mathbf{K}_i\neg p \rightarrow p)\}$ which is equivalent to extrospective representation bel(i,$\neg$K$\neg p \rightarrow p$) has one extension which is included bel(i,p) equivalent to $\mathbf{K}_i p$.
3. Since Jiang's MAE logic is based on Moore's AE logic [Moore 1985] which may not deal with nested belief formulae properly, MAE logic seems to have some undesirable extensions for nested case even in single-agent case. For example, a theory $\{\neg\mathbf{K}\neg\mathbf{K}bel(i, bel(i,p)) \rightarrow bel(i, bel(i,p))\}$ have two extensions: one which has $\neg\mathbf{K}bel(i, bel(i,p))$ and the other has $\mathbf{K}bel(i, bel(i,p))$. Essentially, this phenomenon has happened in Moore's AE logic. Since Jiang's MAE logic extends Moore's AE logic conservatively, the problem is also inherited.

Generalised Multi-agent Autoepistemic Logic

Although our framework is a generalisation of multi-agent autoepistemic reasoning, we shall consider its instantiation which is multi-agent autoepistemic reasoning of Knowledge and can be obtained by instantiating {S4F,SW5} modal systems to our base modal logic.

Generalised proof-theoretic framework

Definition 14 *A Generalised Multi-Agent Autoepistemic (GMAE) theory is defined as a set of GMAE formulae. An S-theory T is a GMAE theory which is closed under the system S.*

Definition 15 *A GMAE theory T' is complete with respect to a sequence of agents $\langle a_1, ..., a_n, k\rangle$ and a GMAE theory T iff for all objective formulae γ in the theory T,*
either $\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_n}\mathbf{K}_k\mathbf{K}_k\gamma \in T'$ or $\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_n}\mathbf{K}_k\neg\mathbf{K}_k\gamma \in T'$.

Example 2
Suppose $T = \{p\}$. Theory T1 $\{\mathbf{K}_i\neg\mathbf{K}_i p, \mathbf{K}_i\neg\mathbf{K}_i\neg p\}$ is complete with respect to agent $\langle i\rangle$ and theory T. Theory T2 $\{\mathbf{K}_j\mathbf{K}_j p, \mathbf{K}_j\neg\mathbf{K}_j\neg p, \mathbf{K}_j\mathbf{K}_j\mathbf{K}_i p, \mathbf{K}_j\neg\mathbf{K}_j\mathbf{K}_i\neg p\}$ is complete with respect to a sequence of agent $\langle j,i\rangle$ and theory T. In addition, theory T3 $\{\mathbf{K}_j\neg\mathbf{K}_j p, \mathbf{K}_j\neg\mathbf{K}_j\neg p, \mathbf{K}_j\neg\mathbf{K}_j\mathbf{K}_i p, \mathbf{K}_j\neg\mathbf{K}_j\mathbf{K}_i\neg p\}$ is also complete respect to a sequence of agent $\langle j,i\rangle$ and theory T.

Definition 16 *A generalised multi-agent nonmonotonic rule of inference* **M** *for a GMAE theory T with respect to a sequence of agents $\langle a_1, ..., a_n\rangle$ is defined as follows:*

$$\frac{T \nvdash \mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_{n-1}}\mathbf{K}_{a_n}\mathbf{K}_{a_n}\phi}{T \vdash \mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_{n-1}}\mathbf{K}_{a_n}\neg\mathbf{K}_{a_n}\phi}$$

Note that this rule allows us to transfer the extrospective form of nonmonotnic reasoning into the scope of modal operators to indirectly handle the introspective form of nonmonotonic reasoning. Indeed, a generalised multi-agent nonmonotonic system $\mathbf{NM}_S$ with respect to a sequence of agents $\langle a_1, ..., a_n\rangle$ is defined as a monotonic modal system S together with this **M** rule of inference.

Definition 17 *A generalised normal proof of ψ with respect to a sequence of agents $\langle a_1, ..., a_i, k\rangle$ from a GMAE theory T in the system NM_S is a sequence of monotonic S-theories T_l which is closed under S monotonic modal system and $0 \le l \le m$ such that*

1. *T_0 is the S-theory generated by T.*
2. *for each $l \le m$,*
   T_{l+1} is the S-theory obtained by adding a formula $\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_i}\neg\mathbf{K}_{a_i}\alpha$ to T_l if
   (a) *$\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_i}\mathbf{K}_{a_i}\alpha \notin T'_l$ where T'_l is the S-theory generated by $T_l \cup \{\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_i}\neg\mathbf{K}_{a_i}\alpha\}$.*
   (b) *for all subformulae $\mathbf{K}_k\phi$ in α, either $\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_i}\mathbf{K}_k\phi \in T_l$ or $\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_i}\neg\mathbf{K}_k\phi \in T_l$.*
3. *$\psi \in T_m$*
   where either $m = 0$ or
   if $m > 0$, then T_m is complete with respect to a sequence of agents $\langle a_1, ..., a_i, k\rangle$ and theory T.

Definition 18 *We define a GMAE extension of a GMAE theory as consisting of those and only those formulae that can be proved by the same sequence of monotonic theories in a generalised normal proof.*

Note that there always exists a GMAE extension for any GMAE theory. We now show how this definition works by some simple examples.

Example 3 *Let T1=$\{\mathbf{K}_i p\}$. Due to the condition 2a and the T axiom of the S modal logic, the formula $\mathbf{K}_j\mathbf{K}_i\neg\mathbf{K}_i p$ cannot be proved. Furthermore, the formula $\mathbf{K}_j\neg\mathbf{K}_i p$ cannot be augmented to the theory because it is not comply with the allowable form of assumption (2). Let T2=$\{\neg\mathbf{K}_i\neg\mathbf{K}_i\mathbf{K}_j p\}$. The formula*

$\mathbf{K}_i\neg\mathbf{K}_i\mathbf{K}_j p$ *cannot be proved from T2. Generally, Let T3=*$\{\neg\mathbf{K}_{a_2}...\mathbf{K}_{a_i}\neg\mathbf{K}_{a_i}p\}$ *and this theory cannot have* $\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_i}\neg\mathbf{K}_{a_i}p$ *as a result.*

Example 4 *Let T4=*$\{\neg\mathbf{K}_i p\}$*. Obviously, there exists a proof of* $\mathbf{K}_i\neg\mathbf{K}_i p$ *and* $\mathbf{K}_j\neg\mathbf{K}_j\mathbf{K}_i p$ *from the theory.*

Example 5 *Let T5=*$\{\mathbf{K}_i(\neg\mathbf{K}_i\neg fly \rightarrow fly)\}$*. This theory has* $\mathbf{K}_i fly$ *as a derivative and* $\neg\mathbf{K}_i fly$ *cannot be proved.*

For the single-agent case where the language is for one agent, we shall consider the representation for single-agent that we discussed briefly in earlier section. The generalised normal proof for single-agent case can be obtained easily by instantiating the above generalised definition with one agent and any modal logic with the N weak necessitation rule of inference ($\phi \vdash \mathbf{K}_i\phi$). As a result, the single-agent normal proof can augment any initial theory with a formula of the form $\neg\mathbf{K}_i\alpha$ where i is the agent name.

Problems revisited and comparison

1. For the sincere assertion causing the problem (2) in Morgenstern's MANML [Morgenstern 1990], our approach can deal with it properly and $\mathbf{K}_h\mathbf{K}_h\mathbf{K}_s p$ is concluded in its GMAE extension. Although $\mathbf{K}_h\mathbf{K}_s(\neg\mathbf{K}_s p \rightarrow \neg\mathbf{K}_s p)$ can be derivable in the theory T in Morgenstern's approach (2), a formula $\mathbf{K}_h\mathbf{K}_s\neg\mathbf{K}_s p$ is not provable due to the condition 2a in the Definition 17. Furthermore, for the insincere assertion producing the problem (2) in LK, our approach however provides an intuitive result where $\mathbf{K}_h\neg\mathbf{K}_h p$ and $\mathbf{K}_h\neg\mathbf{K}_h\mathbf{K}_s\mathbf{K}_h p$ are in its GMAE extension and both $\neg\mathbf{K}_h\neg\mathbf{K}_s\mathbf{K}_h p$ and $\neg\mathbf{K}_h\neg\mathbf{K}_s\mathbf{K}_s\mathbf{K}_h p$ cannot be proved because of the condition 2a in the Definition 17.
2. It can be seen easily that due to our additional condition on *complete* proof in the condition 3 in the Definition 17, the reasoning in taxonomic hierarchies can be captured properly. Unlike Morgenstern's MANML and Parikh's LK that have two extensions for the example in the problem (3) in MANML and (3) also in Parikh discussed previously, our proof yields only one GMAE extension containing fly(tweety).
3. Reasoning about default can be redeemed intuitively in our framework under **S4F** modal system. For instance, consider a theory $\{\mathbf{K}_i(\neg\mathbf{K}_i p \rightarrow q)\}$. $\mathbf{K}_i q$ can be proved but $\mathbf{K}_i p$ cannot be shown and this corresponds to the exact meaning of the default in the theory. Note that this default can be dealt intuitively in our framework only if the proof is based on **S4F** modal system which was proposed to be a modal logic for default reasoning [Truszczynski 1991{a,b}].
4. Consider the problem (3) in Halpern's HM-ONL. HM-ONL still cannot yield a consistent result for the theory $\{\mathbf{K}_i p \vee \mathbf{K}_i q\}$ properly even though other modal logic such as **S4F** is used instead of **KD45**, **K45** modal logic. Due to the sequential proof-steps in our proof theory, this theory can be dealt with properly.
5. In reasoning about default $\{\neg\mathbf{K}_i\neg\phi \rightarrow \phi\}$, the proposed framework does not always derive ϕ depending on whether ϕ is objective (ie. contain no **K** operator) or not. This is why condition 2b in the Definition 17 is used. This condition provides a constructive way of handling nested and self-referential introspection in the same spirit as HAE [Konolige 88a] but without hierarchical levels. In this way, the proposed framework allows an agent to be *sceptical* about other agents' knowledge but *credulous* about other agents' ignorance.

   Consider a theory $\{\mathbf{K}_j(\neg\mathbf{K}_j\neg\mathbf{K}_i fly \rightarrow \mathbf{K}_i fly)\}$, $\mathbf{K}_j\mathbf{K}_i fly$ should not be concluded as far as the *sceptical* notion is concerned. However, $\neg\mathbf{K}_j\mathbf{K}_i fly$ should be derived, instead. Because of the condition 2b in the Definition 17, our approach does not produce $\mathbf{K}_j\mathbf{K}_i fly$ as Morgenstern's MANML and Parikh's LK do. However, consider another theory $\{\mathbf{K}_j(\neg\mathbf{K}_j\neg\mathbf{K}_j fly \rightarrow \mathbf{K}_j fly)\}$. The default rule in this theory is to know possibly about the agent's own knowledge. Since the proposed framework is *credulous* of an agent's ignorance but *sceptical* of its knowledge, our approach therefore does not yield $\mathbf{K}_j fly$. Thus, the seemingly unintuitive extension in Jiang's MAE (3) can be removed within our approach.

The generalisation of Morgenstern's MANML

Similarly to Jiang's MAE logic, our generalised multi-agent autoepistemic framework already captures the notion of Principle of Moderate Arrogance in any depth of nested agents. This is because our generalised multi-agent rule of inference M can bring external introspection of a theory within the scope of modal operators. Together with the employed modal systems which always have the K-axiom, there is no need to define explicitly Principle of Moderate Arrogance. In particular, the rule M implicitly captures all levels (ie. any depth of nested agents) of Principle of Moderate Arrogance.

However, to obtain the Principle of Cautious Arrogance, our framework need be modified in the same way as Morgenstern did for MANML framework.

Definition 19 *A generalised multi-agent nonmonotonic rule of inference M′ for the Principle of Cautious Arrogance for a GMAE theory T with respect to a sequence of agents* $\langle a_1, ..., a_n\rangle$ *is defined as follows:*

$$\frac{T \nvdash \mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_{n-1}}\mathbf{K}_{a_n}\mathbf{K}_{a_n}\alpha \wedge T \nvdash \mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_{n-1}}\alpha}{T \vdash \mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_{n-1}}\mathbf{K}_{a_n}\neg\mathbf{K}_{a_n}\alpha}$$

Indeed, by replacing the rule M with M' in the system NM_S, we capture the Principle of Cautious Arrogance. Also, We can have the generalised normal proof for the Principle of Cautious Arrogance by replacing the condition 2a in Definition 17 with the following condition 2a′:

$\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_{i-1}}\mathbf{K}_{a_i}\mathbf{K}_{a_i}\alpha \notin T'_l$ *and*

$\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_{i-1}}\alpha \notin T'_l$

where T'_l is the S-theory generated by $T_l \cup \{\mathbf{K}_{a_1}\mathbf{K}_{a_2}...\mathbf{K}_{a_i}\neg\mathbf{K}_{a_i}\alpha\}$.

Conclusion

In this paper, we have provided a comprehensive analysis of existing approaches to multi-agent autoepistemic reasoning. We have proposed an integrated theory that deals with the problems of existing approaches. We have shown that this theory is a simple modification of Parikh's LK logic [Parikh 1991]. The theory is general and constructive. Furthermore, it retains the advantages but removes the peculiar results from the existing frameworks. Some principles about multi-agent autoepistemic reasoning such as the principle of Cautious Arrogance, can be embedded naturally in our theory. As a result, our theory generalises Morgenstern's approaches [Morgenstern 1990].

However, in this paper, we took the proof-theoretic point of view to study multi-agent autoepistemic reasoning and our framework is proof-theoretically defined. In our next paper, we shall consider its semantics and other notion of agents reasoning such as credulous reasoning.

Acknowledgements

The first author would like to acknowledge his financial support from the Royal Thai Government Scholarship and especially from the conference to make his trip possible. The second author was an advanced fellow of British Science and Engineering Research Council at Imperial College where the majority of the work is conducted.

References

Appelt, D. and Konolige, K. (1988) A nonmonotonic logic for reasoning about speech acts and belief revision. In the Proceedings of the Second Nonmonotonic reasoning workshop, 164–175:Springer-Verlag.

Finin, T.; Fritzson R.; Mckay D.; McEntire R. (1994) KQML as an Agent Communication Language. In the Proceedings of the Third International Conference on Information and Knowledge Management:ACM press.

Gelfond, M. (1988) Autoepistemic logic and formalization of commonsense reasoning : preliminary report, In the Proceedings of the Second Nonmonotonic reasoning workshop, 176–185:Springer-Verlag.

Genesereth, M. and Fikes, R. et. al. (1992) Knowledge Interchange Format. version 3.0 reference manual, Technical report, Dept. of Computer Science, Stanford Univ.

Genesereth, M. and Ketchpel, S. (1994) Software Agents. *Communication of ACM* 37(7):48–53.

Halpern, J. (1993) Reasoning about only knowing with many agents. In the Proceedings of National Conference on Artificial Intelligence(AAAI-93):655–661.

Halpern, J. and Moses, Y. (1984) Towards a theory of knowledge and ignorance. reprinted in 1985 in Logics and Models of Concurrent system (ed.,K.Apt), 459–476:Springer-Verlag.

Jiang, Y.J. (1994) On multiagent Autoepistemic logic - an extrospective view. In the Proceedings of the International Conference on Knowledge Representation and Reasoning(KR-94).

Konolige, K. (1988a) Hierarchic Autoepistemic Theories for nonmonotonic reasoning: preliminary report. In the Proceedings of the Second Nonmonotonic reasoning workshop, 42–59:Springer-Verlag.

Konolige, K. (1988b) On the relationship between default and Autoepistemic logic. *Artificial Intelligence* 35:343–382.

Labrou, Y. and Finin, T. (1994) A semantics approach for KQML - a general purpose communication language for software agents. In Third International Conference on Information and Knowledge Management:ACM press.

Lakemeyer, G. (1993) All they know: A study in multi-agent Autoepistemic Reasoning. In the Proceedings of the International Joint Conference on Artificial Intelligence(IJCAI-93).

Levesque, H. (1991) All I know: A study in Autoepistemic Logic. *Artificial Intelligence* 42:263–309.

Moore, R. (1985) Semantic considerations of nonmonotonic logic. *Artificial Intelligence* 25:75–94.

Morgenstern, L. (1990) A formal theory of Multiple Agent Nonmonotonic Reasoning. In the Proceedings of National Conference on Artificial Intelligence(AAAI-90):538–544.

Parikh, R. (1991) Monotonic and Nonmonotonic logics of knowledge. *Fundamenta Informaticae* 15(3,4):255–274.

Perrault, R. (1987) An application of default logic to Speech act theory. In Proceedings of the Symposium of Intentions and Plans in Communication and Discourse.

Reiter, R. (1980) A logic of default reasoning. *Artificial Intelligence* 13:81–132.

Truszczynski, M. (1991a) Embedding default logic into modal nonmonotonic logics. In the Proceedings of the First workshop on Logic programming and nonmonotonic reasoning, 151–165:MIT press.

Truszczynski, M. (1991b) Modal Interpretations of Default Logic. In the Proceedings of the International Joint Conference on Artificial Intelligence(IJCAI-91):393–398.