TU Clausthal

Clausthal University of Technology

Communicating Rational Agents: Semantics and Verification

Nils Bulling[1] and Koen V. Hindriks[2]

IfI Technical Report Series **IfI-09-02**

IfI

Department of Informatics
Clausthal University of Technology

Impressum

Publisher: Institut für Informatik, Technische Universität Clausthal
Julius-Albert Str. 4, 38678 Clausthal-Zellerfeld, Germany
Editor of the series: Jürgen Dix
Technical editor: Wojciech Jamroga
Contact: wjamroga@in.tu-clausthal.de
URL: http://www.in.tu-clausthal.de/forschung/technical-reports/
ISSN: 1860-8477

The IfI Review Board

Communicating Rational Agents: Semantics and Verification

Nils Bulling[1] and Koen V. Hindriks[2]

[1]Department of Informatics
Clausthal University of Technology, Germany
bulling@in.tu-clausthal.de

[2] Faculty Electrical Engineering, Mathematics and Computer Science
Delft University of Technology, The Netherlands
k.v.hindriks@tudelft.nl

Abstract

We present a computational semantics of communicative actions for rational agent programming languages. Three indicators are used to differentiate declarative, interrogative and imperative messages which replace the usual labels to identify so-called speech acts. We introduce a multi-agent verification logic based on the computational semantics that facilitates reasoning about communicative actions. Subsequently, this multi-agent logic is embedded into a more expressive modal logic over a run-based semantics. We relate both logics and prove expressivity results. Finally, we show how the modal logic can be used to characterize communicative actions as particular speech acts and allows to verify communication among rational agents.

1 Introduction

We introduce a computational semantics for communicative actions *based on mental models*. A shift is made from the traditional speech-act based labels to the exchange of messages differentiated only by grammatical markers. Three markers corresponding to the three sentence types *declarative*, *interrogative*, and *imperative*, which are also distinguished in natural language grammar, are introduced. We believe the notion of a speech act is best used *descriptively* to characterize message exchanges between agents.

The semantics is designed such that it is particularly easy to integrate it into agent programming languages that facilitate programming with mental models (e.g. [10, 1, 4]), where a mental model consists of the *declarative beliefs and goals* of an agent. To this end, a transition semantics is introduced that provides a recipe for implementing communication between such agents. The motivation for the approach presented thus is pragmatic in the sense that it should provide useful communication primitives for programming agents. In this paper we extend the verification logic for single GOAL agents to a logic that can be used to reason about communicating GOAL agents. We then continue to show that this semantics can be embedded into a modal semantics, similar to that of message-passing systems introduced in [7]. This result can be seen as a conservative extension of the single-agent result presented in [10] to the multi-agent setting introduced here; however, in this paper we left out some operators whose addition is straightforward. The operational semantics does not introduce speech acts, which has been the dominating perspective in agent communication, but the logic can be used to reason about communicating agents and we show that communicated messages can be classified as speech acts by means of this logic.

Section 2 introduces the basic multi-agent model presupposed by both the transition as well as the modal semantics. Section 3 introduces the essentials of an agent programming language on top of which Section 4 adds a computational semantics for communicative actions. We also introduce a logic based on the transition semantics. Section 5 embeds this semantics into a modal semantics. The modal logic is used in Section 6 to characterize message exchanges as speech acts. Section 7 concludes the paper.

2 Preliminaries: The Multi-Agent Model

The multi-agent model, which is based on the (interpreted) systems model from [7], assumes a fixed number of agents with associated *agent names* $\mathcal{A}gt = \{a_1, \ldots, a_n\}$. This assumption is not essential here but simplifies the technical presentation. A *global state* g of a multi-agent system (MAS) is a tuple $\langle l_{a_1}, \ldots, l_{a_n}, l_e\rangle$ with l_{a_i} the local state of agent a_i and l_e the state of the environment. We use g_a to denote the local state of agent a. The non-empty set $G = L_{a_1} \times \ldots \times L_{a_n} \times L_e$ represents all *(global) states*.

In each state an agent a may perform an action, drawn from a set of *actions* Act_a. We use α_i to denote actions and assume that Act_a contains an action Λ which corresponds to agent a performing no action. Without loss of generality, we assume that $Act_a \cap Act_b = \emptyset$ when $a \neq b$. Act denotes the union of the action sets of all agents. As we abstract from the details of action selection here, we associate mappings $P_a : L_a \rightarrow \mathcal{P}(Act_a)$ called *programs* with each agent that map a local state to a *non-empty* set of actions from which the agent may nondeterministically select an action to perform. An *agent* may then be defined as a pair $\langle a, P_a\rangle$ with $a \in \mathcal{A}gt$ and P_a a program. Usually, we identify an agent with its name.

The effects of performing an action are represented by a *transition function* $\tau : G \times Act \rightarrow G$. Actions are assumed to update only the local state of the agent performing it. That is, $\tau(g, \alpha)_a = g_a$ whenever $\alpha \notin Act_a$. An exception to this rule will be made below for communicative actions.

The behavior of a multi-agent system is given by a *run* r which is a mapping $\mathbb{N} \rightarrow G \times Act$. $r_1(i)$ (resp. $r_2(i)$) is used to denote the projection of $r(i)$ onto the first (resp. second) component of $r(i)$. We thus use an interleaving semantics to model the execution of a MAS. This means that for every i we have $r_2(i) \in P_a(g_a)$ for some $a \in \mathcal{A}gt$ where $g_a = (r_1(i))_a$ and $r_1(i + 1) = \tau(r(i))$. Additional constraints such as fairness may be added but are not studied in this paper.

Definition 1 (MAS Model) *A* multi-agent system model $\mathcal{R}$*,* system *for short, is defined as a set of runs.*

3 Programming with Mental Models

In this section we briefly introduce the essential notions used to define a GOAL multi-agent system. Multi-agent GOAL extends the single agent GOAL framework in various ways. A mental state of an agent in a multi-agent system is a more complex entity made up of mental models that are associated with each agent in the system. What we call a mental model here corresponds to a mental state of a GOAL agent in the single agent setting (cf. [10]). Finally, the set of built-in actions is extended with a communication primitive and three markers to indicate the type of sentence that is contained within the message.

GOAL rational agents are programs that derive their choice of action from their beliefs and goals. The GOAL agent programming language provides a framework for programming with *mental models* that consist of an agent's beliefs and goals. Whereas in the single agent setting (see e.g. [6]) a mental model consists of the agent's own beliefs and goals only, in the multi-agent setting, that we consider here, we introduce the notion of a mental state that consists of mental models of other agents as well. The idea is that these mental models are used to (partially) reconstruct the beliefs and goals of another agent, given, for example, the messages received from that agent or observations of other actions performed by that agent. Initially, before any message is received or action is perceived, a mental model of another agent may already contain information about that agent including, for example, beliefs that are considered common knowledge. In Section 4 we show how messages received from another agent may be used to construct a mental model of that agent.

We reserve the label *mental state* for composed entities consisting of multiple mental models, one for each agent in the MAS, and the label *mental model* for pairs of beliefs and goals associated with a particular agent. A mental state of an agent thus is a mapping from agent names to mental models. The beliefs and goals of an agent are declarative sentences which are represented in some underlying knowledge representation technology. Here we assume a propositional language $\mathcal{L}_{PL}$ built from a set of propositional atoms *Atom* and the usual boolean connectives. This language is shared between agents and the vocabulary used to communicate are aligned.[1] $\models_{\mathrm{PL}}$ denotes the usual consequence relation associated with $\mathcal{L}_{PL}$, and the special symbol $\bot \in \mathcal{L}_{PL}$ denotes the false proposition. As is common [4, 10], some additional rationality constraints are imposed on a mental model.

[1] Our primary interest here is to introduce a *computational semantics* for communicating agents and to *characterize message exchanges* between agents in a logic that provides the means to reason about communicating agents. Other issues such as the alignment of vocabularies which may arise in open agent systems are not addressed in this paper.

Definition 2 (Mental Models and Mental States) *A* mental model *is a pair* $\langle \Sigma, \Gamma \rangle$ *with* $\Sigma \subseteq \mathcal{L}_{PL}$ *a* belief *and* $\Gamma \subseteq \mathcal{L}_{PL}$ *a* goal base *which satisfy the following rationality constraints:*

- *The belief base is consistent:* $\Sigma \not\models_{PL} \bot$*;*
- *Individual goals are consistent:* $\forall \gamma \in \Gamma : \gamma \not\models_{PL} \bot$*;*
- *Goals are not believed to be achieved:* $\forall \gamma \in \Gamma : \Sigma \not\models_{PL} \gamma$.

A mental state *is a mapping* m *from* $\mathcal{A}gt$ *to mental models, i.e.* $m(a) = \langle \Sigma, \Gamma \rangle$ *is a mental model for each* $a \in \mathcal{A}gt$*. The set of all mental states is denoted by* $MS(\mathcal{A}gt)$.

The intuition is that a mental state m_a encodes a's beliefs about b's beliefs and goals by mapping agent name b to a mental model $m_a(b) = \langle \Sigma, \Gamma \rangle$ where Σ encodes b's beliefs and Γ encodes b's goals. Agent a's beliefs about agent b do not have to correspond with the *actual* mental state of agent b, i.e. it may be the case that $m_a(b) \neq m_b(b)$ where m_b denotes the mental state of agent b. Agent a's private belief and goal bases are accessed by $m_a(a)$. A mental state thus allows for *second-order beliefs* but not higher-order beliefs as we want to ensure the model is both tractable as well as a useful extension of current agent programming languages. To ensure this we need to make a trade-off between expressivity and computational complexity here. As a consequence, in the model proposed, we cannot have, for example, that an agent believes that another agent either believes ϕ or believes $\neg\phi$.

Mental states are concrete instantiations of the local states of Section 2 and we get $G = MS_{a_1} \times \ldots \times MS_{a_n}$ where MS_{a_i} denotes the set of mental states for agent a_i. Agents need to be able to inspect their mental state and the different mental models part of it. In order to do so we introduce belief ($\mathbf{B}^a$) and goal modalities ($\mathbf{G}^a$) which are annotated with agent names in order to refer to the beliefs and goals, respectively, in the mental model associated with agent a. Formulae of the form $\mathbf{B}^a\phi$ and $\mathbf{G}^a\phi$ are called *mental atoms* which may be combined into composed sentences called *mental state conditions* using the boolean connectives conjunction $\wedge$ and negation $\neg$.

Definition 3 (Mental State Conditions) *The language of* mental state conditions *over* $\mathcal{A}gt$*,* $\mathcal{L}_{MS}(\mathcal{A}gt)$*, is defined by:*

$$\psi ::= \mathbf{B}^a\phi \mid \mathbf{G}^a\phi \mid \neg\psi \mid \psi \wedge \psi$$

where $a \in \mathcal{A}gt$ *and* $\phi \in \mathcal{L}_{PL}$.

The semantics of mental state conditions is defined relative to a mental state. The truth of $\mathbf{B}^b\phi$ in a mental state m_a of agent a denotes that agent a believes that agent b believes ϕ, and, similarly, $\mathbf{G}^b\phi$ denotes that a believes b has goal ϕ. That agent a has these beliefs is left implicit in an agent program as this program provides the context in which to evaluate a mental state condition. In case $a = b$ and $\mathbf{B}^a\phi$ (resp. $\mathbf{G}^a\phi$) is true we simply say that agent a believes (resp. has goal) ϕ.

Definition 4 (Semantics of $\mathcal{L}_{MS}$) *Let* m *be a mental state and* $m(a) = \langle \Sigma_a, \Gamma_a \rangle$*. The* *semantics of $\mathcal{L}_{MS}$ is defined by:*

$$\begin{array}{lll} m \models_{MS} \mathbf{B}^a\phi & \textit{iff} & \Sigma_a \models_{PL} \phi \\ m \models_{MS} \mathbf{G}^a\phi & \textit{iff} & \exists\gamma \in \Gamma_a \textit{ such that } \gamma \models_{PL} \phi \\ m \models_{MS} \neg\psi & \textit{iff} & m \not\models_{MS} \psi \\ m \models_{MS} \psi \wedge \psi' & \textit{iff} & m \models_{MS} \psi \textit{ and } m \models_{MS} \psi' \end{array}$$

A GOAL agent selects actions by means of *action rules*. As the details of action selection are not important in this paper, we abstract away from the specifics and instead use an *agent program* $prg_a : MS_a \to \mathcal{P}(Act)$ which are abstract mappings from mental states to non-empty sets of actions.[2] Then, we may consider a GOAL *agent* as a tuple $\langle a, prg_a \rangle$ where a is an agent name and prg_a an agent program for agent a. We will often identify an agent with his name.

The behavior of an agent is determined by its mental state. The transition functions τ of Section 2 therefore will also be called *mental state transformers*; a corresponding run is also called a(n) *(agent) trace*. As a mental state transformer has type $MS \times Act \to MS$, $\tau(g, \alpha)_a(b)$ must satisfy the constraints from Definition 2 for any $a, b \in \mathcal{A}gt$. Here, $\tau(g, \alpha)$ denotes a global state, $\tau(g, \alpha)_a$ denotes the mental state of agent a, and $\tau(g, \alpha)_a(b)$ denotes the mental model agent a associates with agent b, a notation we will often use below.

4 Communicating Agents

Actions have been defined abstractly as mental state transformers that affect an agent's own mental state. The communicative actions that we introduce here affect the mental state of the receiving agent and may be viewed as complementary to actions that act upon the environment and are useful in a multi-agent context to affect other agents. A communicative action is of the form $send(a, b, msg) \in Act_a$ where msg denotes a message that is being sent by agent a to agent b. This differs from the traditional approach based on speech act theory, where such actions are of the form $send(a, b, label, msg)$ where $label$ identifies the *type* of speech act performed (e.g. *inform*, *query*, etc.; see e.g. [11, 8, 13]) and *msg* denotes the message *content*.[3] The labels available are different for each particular approach (compare e.g. KQML and FIPA). As has been discussed extensively in the literature, one of the main problems, extensively discussed in the literature, concerns the use of speech-act identifying labels as part of the message sent. It is, for example, not possible for the receiving agent to *verify* that the label used to identify the speech act performed corresponds to the act that is actually performed by the sending agent (cf. [15]). For example, agent a may use the label *inform* when sending a message with content ϕ, which requires at least a belief that ϕ is true on the part of agent a according to traditional speech act theory, whereas in fact agent a does

[2] In [7] agent programs correspond to protocols.
[3] In most practical approaches additional parameters are allowed which do not concern us here.

not believe ϕ to be true at all. As agents do not have direct access to the mental states of other agents, it is impossible for agent b to verify such conditions.[4]

To avoid such problems, we do not use speech-act labels but instead use *indicators to identify the sentence type*. Three indicators are introduced that intuitively correspond with the sentence types most often used in natural language: $\bullet$ for *declarative*, **?** for *interrogative*, and **!** for *imperative* sentences. A *message* then is of the form $i\phi$ where i is one of the three indicators.

Definition 5 (Message) *The set $\mathcal{M}sg$ of* messages *consists of all messages of the form $\bullet\phi$, $\mathbf{?}\phi$, or $\mathbf{!}\phi$ where $\phi \in \mathcal{L}_{PL}$.*

Semantically, a shift from the focus on the sender to the hearer is proposed. It is natural that upon receiving a message an agent will update its mental state. The main question then is how? Taking as our example a message $\bullet\phi$ with declarative indicator, different options are available: The receiving agent (i) incorporates ϕ into its own beliefs, (ii) comes to believe that the sender believes ϕ, or (iii) comes to believe the sender intended the receiving agent to believe ϕ. Other, less useful options, are ignored here.[5] We take a definite *engineering stance* here as our main interest is in providing useful communication primitives for programming multi-agent systems which also correspond with basic common sense intuitions. Taking a pragmatic engineering perspective as our starting point, we argue that: option (i) is too strong as it introduces the assumption that the sender always convinces the receiver, and (iii) is too weak as it is no longer very clear what use the communication has and quite complex reasoning patterns would be needed to conclude something useful from such indirect information about the sender's mental state.[6] This leaves option (ii) which, although it may not always be safe, instead only makes the assumption that the sender believes what it says and arguably takes a message at its face value. Obviously, this is not always a safe assumption to make as the sender may be lying. However, it is also not overly presumptuous as the agent just takes the utterance of the sender at face value. This choice also seems reasonable as a *default* way to handle messages from the pragmatic perspective of the well-known Gricean maxims. For example, the Maxim of Quality, which instructs to not say what

[4]The point we are making here does not concern the fact that agents lack certain epistemic abilities. Assuming that agents are autonomous and do not have access to other agents' mental states, it is not realistic to suppose agents would have such capabilities. The point is rather that an agent is not able to *distinguish* between the particular act that is being performed and others; in the example in the main text, the receiving agent without additional knowledge would not be able to distinguish whether the sending agent is, for example, *informing*, *misinforming*, or *lying*.

[5]Of course, there are still other options, one of which is the rather extreme option of *ignoring* all messages, which would make all communication effort redundant. As we are most interested in providing *useful* communication primitives for programming multi-agent systems, the option of ignoring received messages completely is not considered here.

[6]Option (iii) is unsatisfactory as it seems mainly motivated by the idea that an agent should avoid making too strong assumptions at any cost instead of taking a message at its face value. In practice, using this type of semantics as a default would be rather similar to doing nothing with a received message as making good use of information about the intentions of another agent would require rather involved reasoning patterns. In particular, in the context of agent programming, it seems reasonable to conclude that using option (iii) to define a semantics would be too involved to be useful and not worth the effort.

you believe to be false, or for which you lack adequate evidence, may be viewed as supporting the view that taking messages at face value as option (ii) would provide an adequate default to interpret a message without information to the contrary. Finally, the mental models of other agents maintained by each agent and introduced above become particularly useful here again as a means to define the semantics of communication.

We continue to formalize the informal discussion above by extending the mental state transformer τ such that it can be applied to $send(a, b, i\phi)$ actions as well. In order to simplify the presentation, we assume that whenever a message is sent it is immediately received and processed by the recipient.

Definition 6 (Message-Passing MST) *A* message-passing mental state transformer $\tau : G \times Act \to G$ *satisfies:*
• *if* $b \neq a$, $\alpha \in Act_a$, $\alpha \neq send(a, b, \mathsf{m})$*, then* $\tau(g, \alpha)_b = g_b$
• *if* $\alpha = send(a, b, \mathsf{m})$*, then (i)* $\tau(g, \alpha)_i = g_i \; \forall i \in \mathcal{A}gt \setminus \{b\}$*,*
(ii) $\tau(g, \alpha)_b(i) = g_b(i) \; \forall i \in \mathcal{A}gt \setminus \{a\}$*, and*

$$\tau(g, \alpha)_b(a) := \begin{cases} \langle \Sigma_a \oplus \phi, \{\gamma \in \Gamma_a \mid \Sigma_a \oplus \phi \not\models \gamma\}\rangle & \text{if } \mathsf{m} = \bullet\phi \\ \langle \Sigma_a \ominus \phi, \Gamma_a \rangle & \text{if } \mathsf{m} = \mathbf{?}\phi \\ \langle \Sigma_a \ominus \phi, \Gamma_a \cup \{\phi\}\rangle & \text{if } \mathsf{m} = \mathbf{!}\phi \end{cases}$$

Communicating a message m thus modifies the mental model $\langle \Sigma_a, \Gamma_a \rangle$ of the sender a maintained by receiver b as follows:

1. When m is a *declarative*, ϕ is added to the belief base Σ_a to represent the belief of agent b that a believes ϕ. The addition of ϕ to Σ_a is modelled by $\oplus$, where we minimally assume that $\Sigma_a \oplus \phi \models_{\text{PL}} \phi$ when ϕ is consistent. Moreover, we have to remove goals which are satisfied wrt. the updated belief base in order to meet the third rationality constraint from Def. 2.

2. When m is an *interrogative*, ϕ is removed from the belief base Σ_a to represent the belief that agent a does *not* believe ϕ when it asks a question $\mathbf{?}\phi$. The removal of ϕ from Σ_a is modelled by $\ominus$, where we minimally assume that $\Sigma_a \ominus \phi \not\models_{\text{PL}} \phi$ when ϕ is not a tautology.

3. When m is an *imperative*, and ϕ is not a tautology, then it is added to the goal base Γ_a to represent the belief that agent a wants ϕ when it communicates $\mathbf{!}\phi$ and ϕ is removed from the belief base Σ_a to represent the belief that agent a does *not* believe ϕ when it wants ϕ (cf. Definition 2).

The Verification Language $\mathcal{L}_V$. The temporal language $\mathcal{L}_V$ to reason about communicating agents is an extension of the verification logic introduced in [10][7]. In the logic we need to refer to the beliefs an agent has about the beliefs and goals of another agent,

[7] Apart from the global modality $[\alpha]$ referring to action executions which is left out in this paper.

and therefore, a superscript is added to the belief and goal modalities where $\mathbf{B}_a^b\phi$ represents that a believes that b believes ϕ, and, similarly, $\mathbf{G}_a^b\phi$ represents that a believes that b has goal ϕ.

Definition 7 (Verification Language $\mathcal{L}_V$) *$\mathcal{L}_V(\mathcal{A}gt, \mathcal{L}_{PL}, \mathcal{M}sg)$ denotes the set of formulae χ defined by the following grammar:*

$$\chi ::= \textbf{\textit{start}} \mid \mathbf{B}_a^b\phi \mid \mathbf{G}_a^b\phi \mid \neg\chi \mid \chi \wedge \chi \mid \chi \boldsymbol{U} \chi \mid \mathbf{X}\varphi \mid done_a(\alpha)$$

where $\phi \in \mathcal{L}_{PL}$, $\alpha \in Act_a$ and $a, b \in \mathcal{A}gt$. We also write $\mathbf{B}_a$ for $\mathbf{B}_a^a$ and $\mathbf{G}_a$ for $\mathbf{G}_a^a$.

A trace generated by several goal agents and a message passing mental state transformer serves as a model for $\mathcal{L}_V$. Given such a trace and a time point, the semantics of $\mathcal{L}_V$-formulae is defined in a straightforward way.

Definition 8 (Semantics of $\mathcal{L}_V$) *The semantics of $\mathcal{L}_V$-formulae is defined relative to a trace t and a time point $i \in \mathbb{N}$:*

$$\begin{array}{lll}
t, i \models_V \mathbf{B}_a^b\phi & \textit{iff} & g_a \models_{MS} \mathbf{B}^b\phi \textit{ where } g = t_1(i) \\
t, i \models_V \mathbf{G}_a^b\phi & \textit{iff} & g_a \models_{MS} \mathbf{G}^b\phi \textit{ where } g = t_1(i) \\
t, i \models_V \neg\chi & \textit{iff} & t, i \not\models_V \chi \\
t, i \models_V \chi \wedge \chi' & \textit{iff} & t, i \models_V \chi \textit{ and } t, i \models_V \chi' \\
t, i \models_V \mathbf{X}\chi & \textit{iff} & t, i+1 \models_V \chi \\
t, i \models_V \chi \boldsymbol{U} \chi' & \textit{iff} & \exists j \geq i : t, j \models_V \chi' \textit{ and} \\
 & & \forall k : i \leq k < j \Rightarrow t, k \models_V \chi \\
t, i \models_V done_a(\alpha) & \textit{iff} & i > 0 \textit{ and } t_2(i-1) = \alpha
\end{array}$$

5 Embedding $\mathcal{L}_V$ in the Modal Logic $\mathcal{L}_M$

In this section we introduce the modal logic $\mathcal{L}_M$ which is used to reason about runs. Then, we relate the verification logic $\mathcal{L}_V$ and its semantics to the modal logic $\mathcal{L}_M$ and present expressiveness results.

5.1 $\mathcal{L}_M$: Syntax and Semantics

The language $\mathcal{L}_M$ is built from atoms $p \in Atom$ and the temporal constructs $\bigcirc\varphi$ for φ holds in the next state, $\varphi\mathcal{U}\psi$ for φ holds until ψ holds, belief operators $B_a\varphi$ for $a \in \mathcal{A}gt$ believes φ, goal operators $G_a\varphi$ for a has goal φ, and $Done_a(\alpha)$ for a has performed $\alpha \in Act$.

Definition 9 ($\mathcal{L}_M$) *The language $\mathcal{L}_M$ is defined by:*

$$\varphi ::= \mathsf{p} \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_a\varphi \mid G_a\varphi \mid \bigcirc\varphi \mid \varphi\mathcal{U}\psi \mid Done_a(\alpha)$$

The behavior of a MAS is modelled by a set of runs conform Section 2. In addition, an $\mathcal{L}_M$-model consists of the usual belief and goal accessibility relations for each agent, and a valuation which labels states with the facts true in it.

Definition 10 ($\mathcal{L}_M$-Model) *An* $\mathcal{L}_M$-model $\mathfrak{M}$ *is a tuple* $\langle \mathcal{R}, \{\mathcal{B}_a \mid a \in \mathcal{A}gt\}, \{\mathcal{G}_a \mid a \in \mathcal{A}gt\}, V\rangle$ *where* $\mathcal{R}$ *is a set of runs over states* G *and actions* Act, $\mathcal{B}_a, \mathcal{G}_a \subseteq \mathcal{R} \times \mathbb{N} \times \mathcal{R} \times \mathbb{N}$ *are serial belief and goal accessibility relations, respectively, and* $V : \mathcal{R} \times \mathbb{N} \rightarrow \mathcal{P}(Atom)$ *is a valuation function, which assigns to each point the propositional atoms true in it.*

Formulae are interpreted over $\mathcal{L}_M$-models in the standard way (see e.g. [7]). We use $\mathfrak{M}, r, i \models \varphi$ to denote that φ is satisfied on r at time i in model $\mathfrak{M}$.

Definition 11 (Semantics of $\mathcal{L}_M$) *The semantics of* $\mathcal{L}_M$*-formulae over an* $\mathcal{L}_M$*-model* $\mathfrak{M}$*, a run* $r \in \mathcal{R}_{\mathfrak{M}}$*, and a time point* $i \in \mathbb{N}$ *is defined as follows:*

$\mathfrak{M}, r, i \models \mathsf{p}$	*iff*	$\mathsf{p} \in V(r, i)$
$\mathfrak{M}, r, i \models \neg\varphi$	*iff*	$\mathfrak{M}, r, i \not\models \varphi$
$\mathfrak{M}, r, i \models \varphi \wedge \psi$	*iff*	$\mathfrak{M}, r, i \models \varphi$ *and* $\mathfrak{M}, r, i \models \psi$
$\mathfrak{M}, r, i \models B_a\varphi$	*iff*	$\forall (r', i') \in \mathcal{B}_a(r, i) : \mathfrak{M}, r', i' \models \varphi$
$\mathfrak{M}, r, i \models G_a\varphi$	*iff*	$\forall (r', i') \in \mathcal{G}_a(r, i) : \mathfrak{M}, r', i' \models \varphi$
$\mathfrak{M}, r, i \models \bigcirc\varphi$	*iff*	$\mathfrak{M}, r, i+1 \models \varphi$
$\mathfrak{M}, r, i \models \varphi \mathcal{U} \psi$	*iff*	$\exists j : j \geq i$ *and* $\mathfrak{M}, r, j \models \psi$ *s.t.* $\forall k : i \leq k < j \Rightarrow \mathfrak{M}, r, k \models \varphi$
$\mathfrak{M}, r, i \models Done_a(\alpha)$	*iff*	$i > 0$ *and* $r_2(i-1) = \alpha \in Act_a$

We define $X_a(r, i) = \{(r', i') \mid X_a(r, i, r', i')\}$ *for* $X \in \{\mathcal{B}, \mathcal{G}\}$ *and, as usual, abbreviate* $B_a\varphi \wedge \varphi$ *as* $K_a\varphi$.

5.2 Equivalence and Correspondence Results

We formally relate the logics $\mathcal{L}_V$ and $\mathcal{L}_M$ by embedding $\mathcal{L}_V$ into $\mathcal{L}_M$. We do so by introducing a translation tr from $\mathcal{L}_V$-formulae to $\mathcal{L}_M$-formulae and showing that this translation preserves truth. This shows that $\mathcal{L}_M$ can be used to reason about communicating agents instead of the non-standard $\mathcal{L}_V$ (moreover, in a much more expressive way), which is not only useful as standard techniques for modal logic can be applied but also because it allows us to study expressiveness of $\mathcal{L}_V$ compared to $\mathcal{L}_M$.

First, we define the syntactic translation of formulae.

Definition 12 (Translation $tr : \mathcal{L}_V \to \mathcal{L}_M$)

$$
\begin{array}{rcl}
tr(\mathbf{B}_a^b\phi) & = & \begin{cases} B_a B_b \phi & \text{if } a \neq b \\ B_a \phi & \text{if } a = b \end{cases} \\
tr(\mathbf{G}_a^b\phi) & = & \begin{cases} B_a G_b \Diamond \phi & \text{if } a \neq b \\ G_a \Diamond \phi & \text{if } a = b \end{cases} \\
tr(\neg\varphi) & = & \neg tr(\varphi) \\
tr(\varphi \wedge \psi) & = & tr(\varphi) \wedge tr(\psi) \\
tr(\mathbf{X}\varphi) & = & \bigcirc\, tr(\varphi) \\
tr(\varphi \boldsymbol{U} \phi) & = & tr(\varphi)\mathcal{U} tr(\psi) \\
tr(done_a(\alpha)) & = & Done_a(\alpha)
\end{array}
$$

Our first result shows that $\mathcal{L}_M$ and its models are at least as expressive as $\mathcal{L}_V$ over traces; i.e., the modal logic can be used to reason about traces.

Theorem 1 *Let t be a trace. Then there is an $\mathcal{L}_M$-model $\mathfrak{M} = \langle \mathcal{R}, \{\mathcal{B}_a \mid a \in \mathcal{A}gt\}, \{\mathcal{G}_a \mid a \in \mathcal{A}gt\}, V\rangle$ and a run $r^t \in \mathcal{R}$ such that for all $\varphi \in \mathcal{L}_V$ and $i \in \mathbb{N}$ we have: $t, i \models_V \varphi$ iff $\mathfrak{M}, r^t, i \models tr(\varphi)$.*

Proof. (Sketch) Let $\Pi(T) := \{\pi \subseteq Atom \mid \pi \models_{\text{PL}} T\}$ denote the set of all valuations that make $T \subseteq \mathcal{L}_{PL}$ true. The set of runs $\mathcal{R}$ is defined over states $G := L_{a_1} \times \ldots \times L_{a_n} \times L_e$. We set $L_e := \mathcal{P}(Atom)$ and $L_a := MS_a$ for $a \in \mathcal{A}gt$. The run r^t corresponding to trace t is simply defined by: $r_2^t(i) := t_2(i)$, $(r_1^t(i))_e := \emptyset$, and $(r_1^t(i))_a := (t_1(i))_a$ for all $i \in \mathbb{N}$, $a \in \mathcal{A}gt$.

To obtain the belief and goal accessibility relations we define, similar to [10], a set $\mathcal{R}^b$ of all "belief-reachable" runs (peak once runs). That is, $\mathcal{R}^b := \{r \mid \exists! k : (r_1(k) \neq \langle \emptyset, m_\emptyset, \ldots m_\emptyset\rangle), \forall k : r_2(k) = \epsilon\}$ where $\epsilon \notin Act$ and $m_\emptyset$ maps all agent names to $\langle \emptyset, \emptyset\rangle$. The belief accessibility relation for agent a is now defined as follows: $\mathcal{B}_a(r, i, r', i')$ iff (i) $r' \in \mathcal{R}^b$, (ii) $i = i'$, (iii) $r'_e(i) \in \Pi(\Sigma_a)$ where $(r_1(i))_a(a) = \langle \Sigma_a, \Gamma_a\rangle$, and (iv) $\forall b \in \mathcal{A}gt\ ((r'_1(i))_b = (r_1(i))_a)$. That is, a run is $\mathcal{B}_a$-reachable if each local state l_b represents a's mental state (given in the run it is in relation with); and the environment state is used to encode a's beliefs about the world.

Let $\mathcal{R}(\{\gamma_1, \gamma_2, \ldots\}, i) := \{r \mid (r_1(i + j))_e \in \Pi(\{\gamma_j\}), \forall k \in \mathbb{N} \forall a \in \mathcal{A}gt : (r_1(k))_a = m_\emptyset\}$. Now we define the goal accessibility relation as follows: $\mathcal{G}_a(r, i, r', i')$ iff (i) $i = i'$ and (ii) $r' \in \mathcal{R}(\Gamma_a, i)$ where $(r_1(i))_a(a) = (\Sigma_a, \Gamma_a)$. That is, each "goal run" contains one valuation for each goal in Γ_a at some future time point.

Finally, we set $\mathcal{R} := \{r^t\} \cup \mathcal{R}^b \cup \mathcal{R}^g$ where $\mathcal{R}^g := \bigcup_{X \subseteq \mathcal{L}_{PL}, i \in \mathbb{N}} \mathcal{R}(X, i)$ and $V(r, i) := \{\mathsf{p} \in Atom \mid (r_1(i))_e \models_{\text{PL}} \mathsf{p}\}$. That $t, i \models_{\text{V}} \varphi$ iff $\mathfrak{M}, r^t f(t), i \models tr(\varphi)$ is straightforwardly shown by structural induction on φ.

$\varphi = \neg\phi,\ \phi \wedge \psi,\ \mathbf{X}\phi,\ \phi\mathbf{U}\psi,\ done_a(\alpha)$**:** trivial;

$\varphi = \mathbf{B}^b_a\phi,\ a \neq b$**:**

$$
\begin{aligned}
&\mathfrak{M}, r^t, i \models B_a B_b \phi \\
&\text{iff } \forall (r', i) \in \mathcal{B}_a(r^t, i):\ \mathfrak{M}, r', i \models B_b\phi \\
&\text{iff } \forall (r', i) \in \mathcal{B}_a(r^t, i):\ (\forall (r'', i) \in \mathcal{B}_b(r', i): \\
&\qquad \underbrace{\mathfrak{M}, r'', i \models \phi}_{\text{iff } V(r'',i)\models_{\text{PL}}\phi \text{ iff } (r''_1(i))_e\models_{\text{PL}}\phi}\qquad) \\
&\text{iff } \forall (r', i) \in \mathcal{B}_a(r^t, i): \\
&\quad \forall \pi \in \Pi(\Sigma_b),\ (r'_1(i))_b(b) = (\Sigma_b, \Gamma_b),\ \pi \models_{\text{PL}} \phi \\
&\text{iff } \forall \pi \in \Pi(\Sigma_b),\ (r^t_1(i))_a(b) = (\Sigma_b, \Gamma_b),\ \pi \models_{\text{PL}} \phi \\
&\text{iff } \forall \pi \in \Pi(\Sigma_b),\ (t_1(i))_a(b) = (\Sigma_b, \Gamma_b),\ \pi \models_{\text{PL}} \phi \\
&\text{iff } (t_1(i))_a \models_{\text{MS}} \mathbf{B}^b\phi \text{ iff } t, i \models_{\text{V}} \mathbf{B}^b_a\phi
\end{aligned}
$$

$\varphi = \mathbf{G}^a_a\phi$**:**

$$
\begin{aligned}
&\mathfrak{M}, r^t, i \models G_a \Diamond \phi \\
&\text{iff } \forall (r', i) \in \mathcal{G}_a(r^t, i): \exists j \geq i(\mathfrak{M}, r', j \models \phi) \\
&\text{iff } \forall (r', i): (r' \in \mathcal{R}(\Gamma_a, i),\ (r^t_1(i))_a(a) = (\Sigma_a, \Gamma_a) \Rightarrow \\
&\quad \exists j \geq i(\mathfrak{M}, r', j \models \phi)) \\
&\text{iff } \forall (r', i): (r' \in \mathcal{R}(\Gamma_a, i),\ (r^t_1(i))_a(a) = (\Sigma_a, \Gamma_a) \Rightarrow \\
&\quad \exists j \geq i(r'_1(j))_e \models_{\text{PL}} \phi)) \\
&\text{iff } \exists \gamma \in \Gamma_a,\ (r^t_1(i))_a(a) = (\Sigma_a, \Gamma_a), \\
&\quad \forall \pi \in \Pi(\{\gamma\}): (\pi \models_{\text{PL}} \phi) \\
&\text{iff } (t_1(i))_a \models_{\text{MS}} \mathbf{G}^a\phi \\
&\text{iff } t, i \models_{\text{V}} \mathbf{G}_a\phi
\end{aligned}
$$

$\varphi = \mathbf{G}^b_a\phi, \mathbf{B}^a_a\phi$**:** analogously;

■

To obtain a correspondence result in the other direction, it is clear we need to impose some constraints on $\mathcal{L}_M$-models to ensure they model mental states and meet the conditions of Definition 2 and model communicative actions as in Definition 6. Firstly, consider the rationality conditions of Definition 2. The first two consistency conditions are satisfied because of the seriality of the belief and goal relations. An additional postulate is introduced to match the third condition. It is helpful to first define some notation: $\mathcal{B}^b_a(r, i) := (\mathcal{B}_b \circ \mathcal{B}_a)(r, i) = \{(r', i') \mid \exists (r'', i'') \in \mathcal{B}_a(r, i) : (r', i') \in \mathcal{B}_b(r'', i'')\}$; i.e. $\mathcal{B}^b_a(r, i)$ contains all points which are $\mathcal{B}_b$-reachable from some point which is $\mathcal{B}_a$-reachable from (r, i); $\mathcal{G}^b_a := \mathcal{G}_b \circ \mathcal{B}_a$ is defined analogously. To match the third condition, we now introduce the following postulate:

(R1) $\forall a, b \in \mathcal{A}gt : \mathcal{G}_a^b(r,i) \subseteq [\![\Diamond\varphi]\!]_{\mathfrak{M}} \Rightarrow \mathcal{B}_a^b(r,i) \not\subseteq [\![\varphi]\!]_{\mathfrak{M}}$

where $[\![\varphi]\!]_{\mathfrak{M}} := \{(r,i) \mid \mathfrak{M}, r, i \models \varphi\}$, the *denotation of* φ, consists of the points that satisfy φ. The subscript $\mathfrak{M}$ is omited if clear from context.

In order to be able to match the communication semantics of Definition 6, two additional postulates are required. Let r be a run and $X \in \{B, G\}$. The following condition says that only the beliefs and goals of an action executing agent may change provided it is not a send action:

(R2) If $send(\cdot, \cdot, msg) \neq r_2(i) \in Act_a$ then for all $c, d \in \mathcal{A}gt$, $c \neq a$: $X_c(r,i) = X_c(r, i+1)$ and $X_c^d(r,i) = X_c^d(r, i+1)$

Our last postulate handles the case when a message is sent: Only the mental state of the agent who receives the message is allowed to change in a prescribed way.

(R3) If $r_2(i) = send(a, b, msg)$ then for all $c, d \in \mathcal{A}gt$: $X_c(r,i) = X_c(r, i+1)$ and $X_c^d(r,i) = X_c^d(r, i+1)$ except if:

$msg = \bullet\varphi$ and φ consistent then $\mathcal{B}_b^a(r, i+1) \subseteq [\![\varphi]\!]$;

$msg = \mathbf{?}\varphi$ and φ no tautology then $\mathcal{B}_b^a(r, i+1) \not\subseteq [\![\varphi]\!]$;

$msg = \mathbf{!}\varphi$ and φ no tautology then $\mathcal{B}_b^a(r, i+1) \not\subseteq [\![\varphi]\!]$ and $\mathcal{G}_b^a(r, i+1) \subseteq \mathcal{G}_b^a(r,i) \cap [\![\Diamond\varphi]\!]$.

Note that in the case of $\bullet\varphi$, **(R1)** ensures that φ is not a goal. Finally, we call a run *trace-consistent* if it satisfies conditions **(R1)**, **(R2)**, and **(R3)** and an $\mathcal{L}_M$-model is said to be *trace-consistent* if it contains at least one trace-consistent run.

Theorem 2 *Let $\mathfrak{M}$ be a trace-consistent $\mathcal{L}_M$-model. For each trace-consistent run r, all $\varphi \in \mathcal{L}_V$, and $i \in \mathbb{N}$, there is a trace t such that: $\mathfrak{M}, r, i \models tr(\varphi)$ iff $t, i \models \varphi$.*

Proof. (Sketch) Let $\mathfrak{M} = (\mathcal{R}, \{\mathcal{B}_a \mid a \in \mathcal{A}gt\}, \{\mathcal{G}_a \mid a \in \mathcal{A}gt\}, V)$ and $r \in \mathcal{R}$ a trace-consistent run. The trace t is defined from r as follows. For all $i \in \mathbb{N}$ we set (1) $t_2(i) := r_2(i)$ and (2) for all $a, b \in \mathcal{A}gt$ we set $(t_1(i))_a(b) := \langle \Sigma_a^b, \Gamma_a^b \rangle$ where for $a \neq b$: $\Sigma_a^b := \{\phi \in \mathcal{L}_{PL} \mid \mathfrak{M}, r, i \models B_a B_b \phi\}$, $\Gamma_a^b := \{\phi \in \mathcal{L}_{PL} \mid \mathfrak{M}, r, i \models B_a G_b \Diamond\phi\}$; and for $a = b$: $\Sigma_a^a := \{\phi \in \mathcal{L}_{PL} \mid \mathfrak{M}, r, i \models B_a \phi\}$, $\Gamma_a^a := \{\phi \in \mathcal{L}_{PL} \mid \mathfrak{M}, r, i \models G_a \Diamond\phi\}$. The following claims complete the proof.
Claim 1: t is a trace. Proof: That each element of $t_1(i)$ is consistent is ensured by the seriality of the belief relations. Assume some Σ_a is inconsistent. Then there are $\varphi, \psi \in \Sigma_a$ such that $\varphi \wedge \psi$ is inconsistent. Then, $\mathfrak{M}, r, i \models B_a(\varphi \wedge \psi)$ and thus also $\mathfrak{M}, r, i \models B_a\bot$ which contradicts the seriality of $\mathcal{B}_a$. Following the same reasoning, we get that each goal contained in any goal base must be consistent. Finally, that goals are not believed is ensured by postulate **(R1)**. Assume that $\mathfrak{M}, r, i \models G_a \Diamond\gamma \wedge B_a \gamma$, then we would have $\mathcal{B}_a(r,i) \subseteq [\![\gamma]\!]$ and $\mathcal{G}_a(r,i) \subseteq [\![\Diamond\gamma]\!]$, which is a contradiction. It remains to show that t satisfies the conditions imposed by Def. 6. For actions other than communicative actions postulate **(R2)** guarantees that only the mental state of the agent that performs it is affected. For a communicative act, assume that $t_2(i-1) = send(a, b, msg)$ and let $(t_1(i-1))_b(a) = \langle \Sigma_a, \Gamma_a \rangle$ and $(t_1(i))_b(a) = \langle \Sigma'_a, \Gamma'_a \rangle$.

Case $msg = \bullet\psi$**:** For consistent ψ, we show that $\Sigma'_a \models_{PL} \psi$. Assume this is not the case, i.e. $\mathfrak{M}, r, i \not\models B_b B_a \psi$. This is equivalent to $\neg(\forall(r', i') \in \mathcal{B}_b(r, i) : (\forall(r'', i'') \in \mathcal{B}_a(r', i') : (\mathfrak{M}, r'', i'' \models \psi))$ which in turn is equivalent to $\exists(r'', i'') \in \mathcal{B}^a_b(r, i) : \mathfrak{M}, r'', i'' \not\models \psi$. But this contradicts **(R3)**: $\mathcal{B}^a_b(r, i) \subseteq [\![\varphi]\!]$. Moreover, it cannot be the case that there is a $\gamma \in \Gamma'_a$ such that $\Sigma'_a \models_{PL} \gamma$. If that would be the case then $\mathcal{G}^a_b(r, i) \subseteq [\![\Diamond\gamma]\!]$ and $\mathcal{B}^a_b(r, i) \subseteq [\![\gamma]\!]$ which again contradicts (**R1**).

Case $msg = ?\psi$**:** Assume ψ is not a tautology and $\Sigma'_a \models_{MS} \psi$. Then, $\mathfrak{M}, r, i \models B_b B_a \psi$ and hence $\mathcal{B}^b_a(r, i) \subseteq [\![\psi]\!]$, which contradicts **(R3)**.

Case $msg = !\psi$**:** Beliefs are treated as in the previous case. Assume ψ is not a tautology and $\Gamma'_a \not\models \psi$. Then also $\mathfrak{M}, r, i \not\models B_b G_a \Diamond\psi$ which is equivalent to $\neg(\forall(r', i') \in \mathcal{B}_b : \forall(r''.i'') \in \mathcal{G}_a(r', i') : \mathfrak{M}, r'', i'' \not\models \Diamond\psi)$ and hence $\exists(r'', i'') \in \mathcal{G}^a_b(r, i) : \mathfrak{M}, r'', i'' \not\models \Diamond\psi$. So we have that $\mathcal{G}^a_b(r, i) \not\subseteq [\![\Diamond\psi]\!]$ which again contradicts **(R3)**.

Claim 2: $\forall\varphi \in \mathcal{L}_V$: $t, i \models_V \varphi$ iff $\mathfrak{M}, t, i \models tr(\varphi)$
This proof is straightforward by induction on φ.

$\varphi = \mathbf{B}^b_a\phi$**:**

$$\begin{aligned} &t, i \models_V \mathbf{B}^b_a\phi \text{ iff } (t_1(i))_a \models_{MS} \mathbf{B}^b\phi \\ &\text{iff } \{\phi \in \mathcal{L}_{PL} \mid \mathfrak{M}, r, i \models B_a B_b \phi\} \models_{PL} \phi \\ &\text{iff } \mathfrak{M}, r, i \models B_a B_b \phi \end{aligned}$$

$\varphi = \mathbf{G}^b_a\phi$**:**

$$\begin{aligned} &t, i \models_V \mathbf{G}^b_a\phi \text{ iff } (t_1(i))_a \models_{MS} \mathbf{G}^b\phi \\ &\text{iff } \{\phi \in \mathcal{L}_{PL} \mid \mathfrak{M}, r, i \models B_a G_b \Diamond\phi\}\} \models_{PL} \phi \\ &\text{iff } \mathfrak{M}, r, i \models B_a G_b \Diamond\phi \end{aligned}$$

■

The theorems show that both the translation from $\mathcal{L}_V$ to $tr(\mathcal{L}_V)$ preserves truth given that the modal semantics incorporates principles that match those in the computational transition semantics.

The single agent part of the logic $\mathcal{L}_M$ extends the logic discussed in [10], apart from a few operators that were not introduced here due to space limits but which may be added without much effort. The result presented here thus extends that of [10] to the multi-agent case including communicative acts[8]. In particular, this means that we can relate Cohen and Levesque's Intention Logic in a similar way to $\mathcal{L}_V$ as done with $\mathcal{L}_M$ (modulo communication). We do not claim, however, that the logic presented here corresponds to that of [3].

[8] Apart from the $[\alpha]$ modality.

6 Communicative Actions as Speech Acts

The logic $\mathcal{L}_M$ provides us with the means to characterize various communicative actions as instances of particular speech acts. Here, we will focus on such a characterization by means of their associated sincerity and preparatory conditions from standard speech act theory, which are combined into so-called feasibility preconditions in [8].

Remark 3 *A remark about the language used is appropriate as $\mathcal{L}_M$ is strictly more expressive than $tr(\mathcal{L}_V)$. The idea is that we translate a* GOAL *model to an $\mathcal{L}_M$-model to be able to reason about various interpretations of communicative acts from an objective perspective. Hence, it does not make sense to use formulae like $B_a B_b B_c \varphi$ as mental states do not contain related information. Formulae like $B_a(B_b\varphi \vee B_b\neg\varphi) \notin tr(\mathcal{L}_V)$ may be used, however, as well as knowledge operators even though objective truth is not provided by mental states; it is clear, however, that a source for objective truth can be added to a* MAS. *The intuition is that (objective) truth is added to the translated* GOAL *model afterwards. Moreover, it is straightforward how $\mathcal{L}_V$ itself could be extended by a truth layer (e.g. by adding an environment (agent)).*

For example, in [8], the speech act *confirm* has an associated feasibility condition similar to $B_a\phi \wedge B_a\neg B_b\phi$ where a is the sender, b the receiver and ϕ is the message content. Similarly, the so-called *rational effect* of a confirm is informally described as an entitlement of the receiver "to believe that the sender believes the proposition that is the content of the message.", which can be represented as $B_b B_a\phi$.[9]

As a second example from [8], the speech act labeled *inform* has an associated feasibility condition which can be represented by $B_a\phi \wedge \neg B_a(B_b\phi \vee B_b\neg\phi)$[10] whereas the rational effect again may be represented by $B_b B_a\phi$.

Although one can argue about such definitions, e.g. about the difference between a *confirm* and *inform* speech act, it is instructive to take these definitions as a starting point to illustrate how $\mathcal{L}_M$ can be used to characterize a communicative action as a particular speech act. We first characterize a set of speech acts which are all closely related to the *inform* speech act as specified in [8].[11]

Definition 13 (Inform, Misinform, Lie)

$$\begin{aligned}
\mathbf{Inform}_a^b\phi ::= \quad & \bigcirc Done_a(send(a,b,\bullet\phi))\wedge \\
& K_a\phi \wedge \neg B_a Bwh_b\phi \wedge \bigcirc B_b B_a\phi \\
\mathbf{Misinform}_a^b\phi ::= \quad & \bigcirc Done_a(send(a,b,\bullet\phi))\wedge \\
& B_a\phi \wedge \neg\phi \wedge \neg B_a Bwh_b\phi \wedge \bigcirc B_b B_a\phi \\
\mathbf{Lie}_a^b\phi ::= \quad & \bigcirc Done_a(send(a,b,\bullet\phi))\wedge \\
& K_a\neg\phi \wedge \neg B_a Bwh_b\phi \wedge \bigcirc B_b B_a\phi
\end{aligned}$$

[9] [8] also uses an *uncertainty* operator which we have replaced with a negated belief operator. For reasons not clear to us, in [8] the rational effect is represented by $B_b\phi$.

[10] As above, [8] uses the belief modality as well as an uncertainty operator.

[11] $Bwh\phi$ is shorthand for $B_b\phi \vee B_b\neg\phi$; Kwh is similarly defined.

The definitions of the three actions of *informing*, *misinforming* and *lying* only differ with respect to minor but crucial details. We would argue that an agent a *informs* another agent b of a fact ϕ only if agent a actually knows ϕ, i.e. a believes ϕ and ϕ is the case. Similarly, agent a *misinforms* another agent only if a believes ϕ but ϕ is *not* the case. Finally, agent a *lies* about ϕ to another agent only if a knows that $\neg\phi$ but nevertheless communicates a declarative sentence ϕ to that agent. Each of the definitions include a condition that the receiving agent b believes that agent a believes ϕ, which corresponds with the rational effect associated with an inform action in [8] and the default interpretation of the receiving agent of a declarative sentence built into the semantics. Even though it may not be considered rational from a third-person perspective to come to believe that the sender believes ϕ when the sender is lying, as long as an agent does not have information to the contrary, we argued it is a reasonable interpretation strategy to come to believe so upon receiving a declarative sentence.

Typically, only the first act labeled *inform* will be present in speech-act based approaches such as FIPA and there is no room for actions that may be best characterized as misinforming or lying. Although it seems very sensible to try to prevent actions that may be classified in the latter ways, we would argue that it may be hard to ensure this without the proper means to reason about communicative actions. Instead, we argue that using speech acts to reason about communicating agents is more useful than to use speech act labels as part of a communicative message sent to another agent. The work presented is a step towards providing the logical tools to do so in a setting that also provides the means to engineer a multi-agent system, which shows that the logic is computationally grounded.

We provide some additional illustrations related to an agent that is *querying* about a proposition ϕ.

Definition 14 (Query, Exam Query, Rethorical Query)

$$
\begin{array}{rl}
\mathbf{Query}_a^b\phi ::= & \bigcirc Done_a(send(a,b,\mathbf{?}\phi))\wedge \\
& \neg K_a\phi \wedge B_a Kwh_b\phi \wedge \bigcirc B_b\neg B_a\phi \\
\mathbf{ExamQy}_a^b\phi ::= & \bigcirc Done_a(send(a,b,\mathbf{?}\phi))\wedge \\
& Kwh_a\phi \wedge \neg B_a Kwh_b\phi \wedge \bigcirc B_b\neg B_a\phi \\
\mathbf{RethQy}_a^b\phi ::= & \bigcirc Done_a(send(a,b,\mathbf{?}\phi))\wedge \\
& K_a\phi \wedge B_a Kwh_b\phi \wedge \bigcirc B_b\neg B_a\phi
\end{array}
$$

The differences between a proper *query*, an *exam query*, and a *rhetorical query* again are minor but crucial.[12] A message $\mathbf{?}\phi$ is properly labeled a query when the sender does not know ϕ but believes the receiving agent b does know whether ϕ, whereas it may be considered an exam query when the sender knows whether ϕ but does not believe that the receiving agent does too. Finally, a rhetorical query is similar to a proper query except for the fact that the sender knows ϕ.

[12] In [8] the feasibility condition associated with the *query-if* speech act includes a condition that the sending agent does not believe that the receiving agent already intends to inform the sender, a condition we cannot express in the language $\mathcal{L}_V$.

7 Conclusion

We introduced a *computational semantics for agents that communicate at the knowledge level* [12] and a *logic to reason about communicating agents* that may be used to characterize the message exchanges between agents. Messages are used to reconstruct a model of the sender, which also ensures the autonomy of the receiving agent as that agent's beliefs and goals are not directly affected while the mental model of the sender still may be used to further the agent's goals.

[5] seems to shift the burden to implement a speech-act based semantics to the programmer, as they require so-called practical reasoning rules to process received messages. [16] also proposes a shift from sender to receiver in a communication semantics for AgentSpeak(L). The approach is event-driven, however, and does not provide a declarative semantics based on mental models nor a logic to reason about communicative actions.

The expressiveness of the logic to reason about communicating agents is limited compared to other logics that have been proposed [14, 3], and remains an issue for future research, but an advantage of our approach is that it is based on a computational semantics for communication that has typically been missing in pure logical approaches. We believe nevertheless that it is useful to investigate various ways to increase the expressiveness of the language. The recent work presented in [9] on an extension of the GOAL programming language such that a temporal logic can be used as the knowledge representation language may be one interesting option to pursue in future work to do so.

Finally, although in the semantics proposed here agents internalize the meaning of messages, there are interesting links with social commitment semantics [2], and it may be useful to integrate the dynamics of commitments at a social level with our agent-based semantics.

References

[1] L. Braubach, A. Pokahr, W. Lamersdorf, and D. Moldt. Goal Representation for BDI Agent Systems. In *The 2nd International Workshop on Programming Multiagent Systems (ProMAS'04)*, pages 9–20, 2004.

[2] A.K. Chopra and M.P. Singh. Constitutive interoperability. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS'08)*, pages 797–804, 2008.

[3] P.R. Cohen and H.J. Levesque. Communicative actions for artificial agents. In *Proc. of the 1st Int. Conf. on Multi-agent Systems (ICMAS'95)*, 1995.

[4] M. Dastani. 2APL: a practical agent programming language . *Journal Autonomous Agents and Multi-Agent Systems*, 16(3):214–248, 2008.

[5] M. Dastani, J. van der Ham, and F. Dignum. Communication for goal directed agents. In *Communications in MAS*, pages 239–252, 2003.

[6] F.S. de Boer, K.V. Hindriks, W. van der Hoek, and J.-J.Ch. Meyer. A Verification Framework for Agent Programming with Declarative Goals. *Journal of Applied Logic*, 5(2):277–302, 2007.

[7] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT, 1995.

[8] FIPA. Fipa communicative act library specification. Technical Report SC00037J, Foundation for Intelligent Physical Agents, Geneva, Switzerland, 2002.

[9] K. Hindriks, M.B. van Riemsdijk, and W. van der Hoek. Agent programming with temporally extended goals. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'09)*, 2009. To appear.

[10] K.V. Hindriks and W. van der Hoek. GOAL agents instantiate intention logic. In *Proc. of the 11th European Conference on Logics in Artificial Intelligence (JELIA'08)*, pages 232–244, 2008.

[11] Y. Labrou and T. Finin. A semantics approach for KQML - a general purpose communication language for software agents. In *3rd Int. Conf. on Information and Knowledge Management (CIKM'94)*. 1994.

[12] A. Newell. The Knowledge Level. *Artificial Intelligence*, 18(1):87–127, 1982.

[13] J.R. Searle. *Speech acts*. Cambridge University Press, 1969.

[14] M.P. Singh. *Multiagent Systems: A Theoretical Framework for Intentions, Know-How, and Communications*. Springer-Verlag, 1994.

[15] M.P. Singh. Agent Communication Languages: Rethinking the Principles. *IEEE Computer*, 31(12):40–47, 1998.

[16] R. Vieira, A. Moreira, M. Wooldridge, and R.H. Bordini. Formal Semantics of Speech-Act Based Communication in an Agent-Oriented Programming Language. *Artificial Intelligence Research*, 29:221–267, 2007.