

A Combined Logic of Expectation and Observation

A generalisation of BDI logics

Bình Vũ Trần, James Harland, and Margaret Hamilton

School of Computer Science and Information Technology
RMIT University, Australia
{tvubinh, jah, mh}@cs.rmit.edu.au

Abstract. Although BDI logics have shown many advantages in modelling agent systems, the crucial problem of having computationally ungrounded semantics poses big challenges when extending the theories to multi-agent systems in an interactive, dynamic environment. The root cause lies at the *inability of modal languages to refer* to the world states which hampers agent reasoning about the connection of its mental attitudes and its world. In this paper, following ideas in hybrid logics, we attempt to readdress the computational grounding problem. Then, we provide a formalism for observations – the only connection between mind and worlds – and expectations – the mental states associated with observations. Finally, we compare our framework with BDI logics.

1 Introduction

The most widely held view for practical reasoning agents is that they are *intentional systems* whose behaviour can be explained and predicted through the attribution of mental attitudes such as beliefs, desires, hopes, fears. . . Since the seminal work of Hintikka [14], formal analyses of mental attitudes are mainly carried out using *modal logics*. Among these models, Belief-Desire-Intention (BDI) model [7] and BDI logics [18] have been one of the most successful. Unfortunately, BDI logics are usually claimed as having *ungrounded semantics* [21], that is, there have been no work showing a one-to-one correspondence between mental models and any concrete computational interpretation. This results in a large gap between theory and practice [16]. The problem, however in our view, should be stated more precisely that since the relationship between mental and computational models realizing the same modal language is a many-to-many relation, it is unclear which computational model is the most suitable for simulating an agent’s mental model in a dynamic interactive environment.

In this paper, we will demonstrate that the inability to find a concrete and useful computational model is due to the lack of expressive power in modal languages which are used as agent specifications. Since modal syntax offers no grip on worlds, from the local, internal perspective of modal logics, it is impossible to detect any difference between the structures of two models. Consequently,

agent reasoning capabilities are seriously impaired. For example, imagine an eagle is chasing a sparrow in a cave system where every cave appears identical. At any cave, there is only one identical unidirectional passageway to another cave. Modal language would express this by saying “All caves accessible from this cave are identical to it.” But the eagle cannot tell whether it has flown through a cave before. So either the cave system has an infinite number of caves connected with each other, so that the eagle cannot visit a cave twice, or it has only one cave looping back on itself, the eagle would not be able to distinguish using orthodox modal language. If a distinction could be recognised, the eagle would be able to justify its expectation where the sparrow could be. Hence it would speed up to catch the sparrow in the former case, but it would stay still in the current cave waiting for the sparrow in the latter case.

It may become apparent that a mechanism to mark the visited worlds will be a significant advantage for the exploring agent allowing it to redraw a map of the real world in its mind. It would not only help the agent to differentiate between two different models, but also provide a tool to base its future predictions. Hybrid languages by Blackburn and Tzakova [6], [3] provide such a naming mechanism for modal languages by introducing a unique label of each world and an operator to jump and evaluate formulae in any world. We believe that such mechanism is strongly related to the concept of observation – the only connection between mind and world. Hence, a formalism that describes observation and its associated mental states, expectations will bridge the mental models and the computational models.

In this paper, following Blackburn and Tzakova we develop a formalism for expectation and observation in hybrid logics and compare this with the BDI model. The paper commences by elaborating in detail the computational grounding problem, and giving an overview of hybrid logics. We then describe the observation-expectation system’s details in section §3 and its comparison with BDI logics in section §4. Finally, we briefly outline our approach towards the application of our framework in multi-agent systems.

2 Computational grounding problem

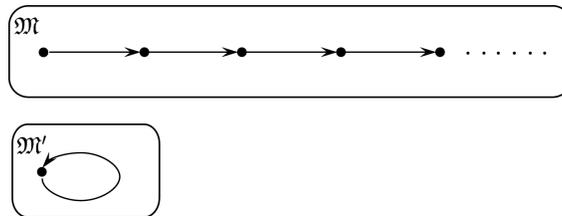


Fig. 1. Invariance and modal languages

The principal motivation for modal languages is to provide a local, internal perspective about world structures. The truth value of any formula is evalu-

ated inside the structure, at a particular (current) state. Even though modal operators also provide access to information at other states, only states that are directly accessible from the current are allowed. This property has attracted many scholars in various disciplines such as cognitive science, psychology, and artificial intelligence to use modal logic as a formal analytical tool for mental models.

The relationship between models in modal logic has been well studied under the notion of *bisimulation* [5]. It is revealed that with only restrictions on identical atomic information and matching accessibility relations to make modal languages invariant, bisimulations are *many-to-many* relations.

Definition 1 (Bisimulation) *Let $\Phi = \{p, q, \dots\}$ be a set of atomic propositions. Given two models $\mathfrak{M} = (W, \sim, \pi)$ and $\mathfrak{M}' = (W', \sim', \pi')$, where W, W' are non-empty sets of possible worlds, $\sim \subseteq W \times W, \sim' \subseteq W' \times W'$ are two accessibility relations on W and W' , and $\pi : \Phi \rightarrow \text{Pow}(W), \pi' : \Phi \rightarrow \text{Pow}(W')$ are two interpretation functions which respectively tell the sets of worlds of W and W' where each proposition holds.*

A bisimulation between two models \mathfrak{M} and \mathfrak{M}' is a non-empty binary relation $Z \subseteq W \times W'$ (\mathfrak{M} and \mathfrak{M}' are called bisimilar) if the following conditions are satisfied:

- (**prop**) if wZw' and $w \in \pi(p)$, then $w' \in \pi'(p)$ for all $p \in \Phi$
- (**forth**) if wZw' and $w \sim v$, then there exists $v' \in W'$ such that vZv' and $w' \sim' v'$
- (**back**) if wZw' and $w' \sim' v'$, then there exists $v \in W$ such that vZv' and $w \sim v$

Therefore, two models with completely different structures can be modally equivalent. Given an arbitrary model, there can be many other models bisimilar to it. For example, let's compare a model \mathfrak{M} , which has the natural numbers in their usual order as its frame ($W = \mathbb{N}$) and every propositional symbol is **true** at every world, with another model \mathfrak{M}' which has a single reflexive world as its frame and all propositional symbols are **true** at this world (see **Fig. 1**). Apparently, one is infinite and irreflexive whilst the other is finite and reflexive. However, they both recognise the same modal language.

One may argue that this would be an advantage of modal language. So for example if the model \mathfrak{M} above is a mental model, its identical structure would not be implementable on a computer due to the infinite set of mental states. However, the equivalent model \mathfrak{M}' has a finite set of states, and hence would certainly be implementable. This argument is valid with the assumption that the original mental model is unchanged and fixed.

Unfortunately, we also have two crucial disadvantages, namely, the **inability to model agents in a dynamic environment** and the **verification** problems. Firstly, in a dynamic unpredictable environment, the agent is continuously updating its mental models. The properties of the models may very well be changed under such updates. The simple computational model \mathfrak{M}' must also be updated to simulate exactly its mental model. However, for example, if we add one more

world v related to a particular world w in the above model \mathfrak{M} , where only some propositional symbols are true, it may be clear that a single simple addition to \mathfrak{M}' to reflect the change is not easy. One may keep arguing that perhaps \mathfrak{M}' was not the right choice. In an environment, where change is unpredictable an identical structure always guarantees the number of changes for the computational model is not greater than the number of changes in the mental model. For other non-identical structures, the possibility of more substantial changes does exist.

Secondly, whilst *axiomatic* verification is as hard as the complexity of a proof problem, *semantic* verification (model checking) is more efficient [22, p 296]. Unfortunately, semantic approaches have a serious problem: it is unclear how to derive appropriate accessibility relations from a given *arbitrary* concrete program. It may be clear how to use the relation between computational states of the computational model to construct a unimodal logic. However, for multi-modal logics such as BDI logics, the construction is usually ad hoc. There are two reasons for this difficulty.

1. *Partitioning problem*: It is unclear how to partition the relational structure between computational states, and use these partitions for their corresponding accessibility relations' constructions. It is as hard as an exhaustive search for all possible combinations.
2. *Interaction problem*: Assuming the first problem were solved, for any sub-model of the original computational model \mathfrak{M}' , since bisimulation is a many-to-many relation, there can be many mental models with very different structures corresponding to it. It is also unclear how constraints between modalities can be derived from the sub-models.

The partitioning problem could be reduced if the relational structure can be classified at a meta level, i.e., if it is possible to classify formulae derived from the single relational structure into different sorts. Each sort corresponds to a mental attitude. The partition that realises that sort can then be used to derive the corresponding accessibility relation of the mental modality. This ultimately provides a two-tier approach. The first tier is totally based on a single accessibility relation with an additional sorting mechanism. The second tier with various accessibility relations is then constructed on the first tier. Unfortunately, such sorting mechanism is not provided in orthodox modal languages.

The above issues lead us to a definite conclusion: modal languages are not expressive enough for specifying agents in general and particularly in a dynamic, unpredictable environment. We are unable to update our computational model reliably and cheaply for changes in theory. More seriously, it is also expensive to verify correctness of a computational model. We call this problem *the computational grounding problem of modal languages* (cf. [21]). In other words, modal languages cannot describe the connection between minds and worlds.

Hybrid languages [3], [6], overcome the problem by two simple additions to modal languages: a new sort of formulae *nominals* and *satisfaction operators* @. Basically, nominals are just atomic propositions disjoint from the set of normal propositions. The crucial difference is that, each nominal is **true** only at a unique world in the possible worlds structure. Therefore, nominals can be considered

as the name, or label for worlds. Satisfaction operators @ are used to assert satisfaction of a formula at a certain world. They allow us to jump to the world where the formula is evaluated.

The additions are relatively simple, but they have significant contributions. Firstly, though hybrid bisimulation could be altered slightly by adding nominals to the set of proposition Φ in the (**prop**) rule, the addition of the (@) rule insists that nominals must be true at a unique world in each model.

- (@) for all nominals s , if $\pi(s) = \{w\}$ and $\pi'(s) = \{w'\}$, then wZw' where $w \in W, w' \in W'$.

The (@) rule guarantees every world $w \in W$ has a unique corresponding world $w' \in W'$. Therefore, though bisimulation Z is a relation, under this condition, it becomes a one-to-one function. In other words, a hybrid bisimulation is equivalent to an isomorphism [1]. With these results, two necessary conditions to solve the grounding problem are satisfied.

Secondly, in [1], it was also proved that the addition of nominals and satisfaction operators does not raise complexity. The satisfiability problem remains decidable in *PSPACE-complete*.

The path of hybrid logics that this paper follows is led by Blackburn [2]. Computational complexity and characterisation of hybrid logics were studied in [1]. The web site <http://www.hylo.net> provides further resources and development in hybrid logics.

3 Observation-expectation system

3.1 Agents, observations and expectations

The real world has its own structure and properties. An agent's mental model is only a reflected part of that structure in the agent's mind. The only means that the agent has to discover its environment is through its *observation*. Therefore, in order to formalise the connection between an agent's mind and its real world, we make observation our essential concept.

According to the *Merriam-Webster Unabridged* dictionary [13], an observation is "an act of recognising and noting a fact or occurrence..." or "an act of seeing or of fixing the mind upon anything." In our framework, observation is a bisimulation between the real world and an agent's mental model. A single object in the real world, e.g. the planet *Venus*, can have multiple images in an agent's mind, '*the morning star*' and '*the evening star*' through two different observations. Conversely, a single mental state can refer to various real world objects, e.g. a tiger refers to any individual which is a large carnivorous feline mammal having a tawny coat with transverse black stripes. This ability of rational agents is usually known as *abstraction*.

Unfortunately, this only connection to the real world is not always available due to various reasons, e.g. limitations of sensors, noises or disruptions from the environment. In such conditions, a rational agent is still able to continually

construct the world model in its mind using its inferential mechanisms and act upon this model accordingly. Thus in the above example, when chasing the sparrow, if the sparrow disappears into a passageway, the eagle would predict the sparrow's movement and keep flying to the other end of the passageway to catch the sparrow there, instead of stopping the chase. The images of the world in the agent's mind that are associated with observations as about to happen are called *expectations*.

In this section, firstly we describe how observations are linked in an observation system and the association of observation with expectation. Secondly, we introduce the formalism for expectation logic based on the observation system.

3.2 Observation system

We can now formally define an observation system with its corresponding real-world global states:

Definition 2 (Observation system) *Let \mathcal{I} be a set of agent identities, L_i be a set of local states for any agent a_i , where $i \in \mathcal{I}$. An observation system is a quadruple $OS = \langle \mathbb{G}, \tau, \mathcal{O}, g_0 \rangle$ where*

- $\mathbb{G} \subseteq \prod_{i \in \mathcal{I}} L_i$ is the set of global states with each $g \in \mathbb{G}$ being an instantaneous global state.
- $g_0 \in \mathbb{G}$ is the initial state of the system.
- The environment of an agent a_i is $Env_i \subseteq \prod_{j \in \mathcal{I} \setminus i} L_j$
- An agent a_i 's collection of observations is a relation $Obs^i \subseteq \mathbb{G} \times \mathcal{E}_i$, where $\mathcal{E}_i \subseteq L_i$ is the set of expectations and each pair (g, ε) is called an observation taken by the agent a_i . ε is the agent's expectation about g through the observation.
- Let $\mathfrak{G} \subseteq \mathcal{I}$. A group \mathfrak{G} 's collection of observations is a relation $Obs^{\mathfrak{G}} \subseteq \mathbb{G} \times \mathcal{E}_{\mathfrak{G}}$ where $\mathcal{E}_{\mathfrak{G}} = \bigcup_{i \in \mathfrak{G}} \mathcal{E}_i$. The pair $(g, \varepsilon) \in Obs^{\mathfrak{G}}$ is called a group observation.
- $\tau : \mathbb{G} \times \mathcal{O} \rightarrow \mathbb{G}$ is a system state transformer function that depicts how a global state transits to another when a set of observation methods from the observation method family \mathcal{O} defined below are carried out by some or all agents in the system.

Given a sensor set ($\mathbb{S} = \bigcup_{i \in \mathcal{I}} \mathbb{S}_i$) and a effector set ($\mathbb{E} = \bigcup_{i \in \mathcal{I}} \mathbb{E}_i$), where \mathbb{S}_i and \mathbb{E}_i are respectively the sets of sensors and effectors of an individual agent a_i . Let's look at the eagle and the cave system in the above example as an agent in an observation system. The eagle's sensors \mathbb{S}_i (eyes, ears, skin, ...) and effectors \mathbb{E}_i (wings, neck, ...) and any combination of them bring different observations (Obs_i) about the environment to the eagle. The eyes ($\xi \in \mathbb{S}_i$) bring visual images of the caves ($\varepsilon_1 \in \xi$), and sparrow ($\varepsilon_2 \in \xi$) to the eagle's brain $\mathcal{E}_i \subseteq L_i$. The wings ($e \in \mathbb{E}_i$) when flapping may bring an observation that its position would be closer to the sparrow ε_3 . However, this can be verified by the eagle's eyes if $\varepsilon_3 \in \xi$.

In any observation system, sensors and effectors are the primary sources that generate observations. Each sensor or effector is associated with a set of observations. An important note is that observations associated with an effector are only *hypothetical*. That is, the agent is always *uncertain* about the consequences of its actions until it uses its sensors to verify the results. Thus, an observation of an effector is justified if and only if it is also associated with a sensor. Observing the world by obtaining observations directly from sensors and effectors is called *primitive observation method* \mathbb{O}_0 (e.g. $e, \xi \in \mathbb{O}_0$). A more complicated set of observation methods \mathbb{O}_k would arrange the k results (expectations) of other observation methods in a systematic way to generate new expectations about the world. These expectations are also associated with global states to form more complex observations. Observation methods are formally defined as follows:

Definition 3 (Observation methods) *An observation method family is a set of observation method sets $\mathbb{O} = \{\mathbb{O}_k\}_{k \in \mathbb{N}}$ where \mathbb{O}_k is a set of observation methods of arity k for every $k \in \mathbb{N}^+$. $\mathbb{O}_0 = \mathbb{S} \cup \mathbb{E}$ is called primitive observation method set. \mathbb{O}_k is inductively defined as follows:*

- $\varepsilon \in \mathcal{E}$ for all $\varepsilon \in o_0, \forall o_0 \in \mathbb{O}_0$
- $o_k(\varepsilon_1, \dots, \varepsilon_k) \subseteq \text{Pow}(\mathcal{E})$ for all $o_k \in \mathbb{O}_k$ and $\varepsilon_1, \dots, \varepsilon_k \in \mathcal{E}$

Thus, if the eagle expects the sparrow would reach the end of the passageway, and it also expects with a flap it would get to the same place at the same time, a combination of the two expectations $o_2(\varepsilon_2, \varepsilon_3) \in \mathbb{O}_2$ provides a way of chasing the sparrow which generates an expectation that two birds would be at the same place (ε_4). However, if the eagle executed the “chasing the sparrow” method $o_2(\varepsilon_2, \varepsilon_3)$, and it got the expectation ε_5 that the sparrow was not at the same place, it would not be able to distinguish the two resulting expectations ε_4 and ε_5 .

Intuitively, whilst adopting any observation method o_k , an agent may have various observations about a real world state g which bring various expectations to the agent’s mind. We call these expectations *indistinguishable* to the observation method o_k about the world g . That is, we cannot tell by adopting observation method o_k what makes the generated observations different. Also, two sets of a group of agents’ expectations will appear indistinguishable to an individual agent if its expectations in both sets do not change. Formally,

Definition 4 (Indistinguishability and transparency properties)

- Two sets of expectations $\mathcal{E}, \mathcal{E}'$ are indistinguishable through an observation o_k if $\mathcal{E}, \mathcal{E}' \in o_k(\varepsilon_1, \dots, \varepsilon_k)$.
- Two sets of expectations $\mathcal{E} = \{\varepsilon_1, \dots, \varepsilon_k\}, \mathcal{E}' = \{\varepsilon_1, \dots, \varepsilon_j\}$ are indistinguishable to an agent a_i if $l_i(\mathcal{E}) = l_i(\mathcal{E}')$ where $l_i : \text{Pow}(\mathcal{E}) \rightarrow \text{Pow}(L_i)$ is a function that extracts the local states of an agent from a set of expectations.
- An observation method o_k is considered as transparent if $o_k(\varepsilon_1, \dots, \varepsilon_k) = \{\mathcal{E}\}$ where $\mathcal{E} = \{\varepsilon_1, \dots, \varepsilon_k\}$ (cf. [23]). A transparent observation method can also be considered as a compound sensor. The family of transparent observation methods is denoted as \mathbb{O}^t .

Indistinguishability of observation methods unfortunately also brings uncertainty to agents. Thus, if $o_2(\varepsilon_2, \varepsilon_3) = \{\{\varepsilon_4\}, \{\varepsilon_5\}\}$, the eagle would be unsure which result would occur. Transparent observation methods hence are more preferable. However, that also means the eagle must determine a correct set of observations and take sufficient observations for a transparent observation method. The eagle could combine other observations such as no way out of the cave, or no other eagle close to the sparrow to guarantee that its “chasing the sparrow” output could be more certain.

3.3 Expectation logic

We can now study how an agent generates its expectations from its observations about its environment by introducing an expectation logic \mathcal{L} based on an observation system. Consider the set of agents identified by the identity set \mathcal{I} . To view an observation system as a hybrid Kripke structure, we introduce the observation interpretation function $\pi : (\Phi \cup \Xi) \rightarrow Pow(\mathbb{G})$ which based on available set of observations to tell which expectation is associated with which global state. $\Phi = \{p, q, r, \dots\}$ is called the primitive expectation proposition set and $\Xi = \{s, t, \dots\}$ is called the observation naming set. The crucial difference from orthodox modal logic is that for every observation name s , π returns a singleton. In other words, s is **true** at a unique global state, and therefore tags this state. Yet, it is possible that a global state can have different observation names. We refer the couple $\mathfrak{M} = \langle OS, \pi \rangle$ as a model of expectation in an observation system.

Definition 5 *The semantics of expectation logic \mathcal{L} are defined via the satisfaction relation \models as follows*

1. $\langle \mathfrak{M}, g \rangle \models p$ iff $g \in \pi(p)$ (for all $p \in \Phi$)
2. $\langle \mathfrak{M}, g \rangle \models \neg\varphi$ iff $\langle \mathfrak{M}, g \rangle \not\models \varphi$
3. $\langle \mathfrak{M}, g \rangle \models \varphi \vee \psi$ iff $\langle \mathfrak{M}, g \rangle \models \varphi$ or $\langle \mathfrak{M}, g \rangle \models \psi$
4. $\langle \mathfrak{M}, g \rangle \models \varphi \wedge \psi$ iff $\langle \mathfrak{M}, g \rangle \models \varphi$ and $\langle \mathfrak{M}, g \rangle \models \psi$
5. $\langle \mathfrak{M}, g \rangle \models \varphi \Rightarrow \psi$ iff $\langle \mathfrak{M}, g \rangle \not\models \varphi$ or $\langle \mathfrak{M}, g \rangle \models \psi$
6. $\langle \mathfrak{M}, g \rangle \models \langle \mathcal{E}_i \rangle \varphi$ iff $\langle \mathfrak{M}, g' \rangle \models \varphi$ for some g' such that $g \sim_e^i g'$
7. $\langle \mathfrak{M}, g \rangle \models [\mathcal{E}_i] \varphi$ iff $\langle \mathfrak{M}, g' \rangle \models \varphi$ for all g' such that $g \sim_e^i g'$
8. $\langle \mathfrak{M}, g \rangle \models s$ iff $\pi(s) = \{g\}$ (for all $s \in \Xi$), g is called the denotation of s
9. $\langle \mathfrak{M}, g \rangle \models @_s \varphi$ iff $\langle \mathfrak{M}, g_s \rangle \models \varphi$ where g_s is the denotation of s .

where 1 – 7 are standard in modal logics with two additions of hybrid logics in 8 and 9.

We have introduced the modality \mathcal{E}_i which allows us to represent the information of the environment resident in the agent a_i 's mind about the output of its observation methods. The semantics of the \mathcal{E}_i modality are given through the *expectation accessibility relation* defined as follows

Definition 6 Given a binary expectation accessibility relation $\sim_e^i \subseteq \mathbb{G} \times \mathbb{G}$ then $g \sim_e^i g'$ iff $\exists \mathcal{E} \subseteq g, \exists \mathcal{E}' \subseteq g'$, such that $\mathcal{E}, \mathcal{E}'$ are indistinguishable to the agent a_i through an arbitrary existing observation method $o_k(\varepsilon_1, \dots, \varepsilon_k)$, where $\varepsilon_1, \dots, \varepsilon_k \in g$ and $\varepsilon_1, \dots, \varepsilon_k \in g'$.

Thus, if $[\mathcal{E}_i]\varphi$ is true in some state $g \in \mathbb{G}$, then by adopting observation method o_k , the local states of the agent a_i about the environment (its expectations) remains the same. An eagle expects to catch a sparrow in a cave, if and only if *wherever* it adopts the observation method “chasing the sparrow” above, the sparrow will appear close to it.

The last two lines in the **Definition 5**. hybridise expectation language \mathcal{L} . Satisfaction operator $@_s$ is considered as an observation operator. Thus, a formula such as $@_s\varphi$ says there is an observation about the world state labelled as s which makes φ **true** in the agent’s mind. For example, by assigning the current cave to s , and “seeing the sparrow” to $p (= \varepsilon_2)$, $@_sp$ tells us the sentence “I am seeing a sparrow in cave s .” This formula remains valid even though the eagle’s current cave is no longer s .

Observation operators are in fact *normal modal operators* (i.e. $@_s(\varphi \Rightarrow \psi) \Rightarrow (@_s\varphi \Rightarrow @_s\psi)$). However, specially it is a *self-dual operator* ($@_s\varphi \Leftrightarrow \neg @_s\neg\varphi$). This can be read as *for all* observations about s , φ holds in the agent’s mind if and only if *there exists* no observation about s that brings $\neg\varphi$ to its mind. We can use ‘*for all*’ and ‘*there exists*’ in this sentence interchangeably. It also allows the expression of state equality “In cave s , it is also named Happy Cave”, by $@_su$ if u represents ‘Happy Cave’.

A formula with both observation operator and expectation modality is more interesting. $@_s\langle\mathcal{E}_i\rangle t$ will tell us that the eagle expects one of the next caves from s will be t . In other words, there is an observation about the connection from the cave s to the cave t . $@_s[\mathcal{E}_i]t$ strongly asserts that t is the only subsequent cave the eagle expects (since t is true at a unique world).

3.4 Expectation reasoning – Observation logic

The construction of expectation logic is strongly dependent on two crucial factors: the set of observations, which provides a basis to observation interpretation function π for assigning truth values to formulae, and the set of observation methods which is the skeleton for constructing accessibility relation between expectations. Unfortunately, due to limitations of primitive sensors and effectors, an agent will not always be able to obtain all observations about the real world.

Therefore, in such conditions a rational agent should carefully select its observations in order to maximise the synchronisation between its mental models and the real world. Thus, when the sparrow flies into a dark passageway, the eagle can no longer take an observation of its prey. Yet, based on its existing expectations (mental images) and available observation methods at the current world, the eagle can still deliberate and determine the next observation to take. Such deliberation is possible since the eagle’s reasoning now relies on another model – the *attention model*.

Definition 7 (Attention model) A model of the mental states at world s is called an attention model $\mathcal{A}_i(s) = \langle W_e, \sim_{\mathbb{O}_s}^i, \rho_s \rangle$ where W_e is the set of expectation worlds which are uniquely named by primitive propositions $p \in \Phi$, the function $\rho_s : \Xi \rightarrow \text{Pow}(W_e)$ interprets what are possible expectations for an observation at s , and $\sim_{\mathbb{O}_s}^i \subseteq W_e \times W_e$ is an observability accessibility relation.

The semantics of observation logic are defined via the satisfaction relation \models_s as follows

- $\langle \mathcal{A}_i(s), w \rangle \models_s t$ iff $w \in \rho_s(t)$ (for all $t \in \Xi$)
- $\langle \mathcal{A}_i(s), w \rangle \models_s p$ iff $\rho_s(p) = \{w\}$ (for all $p \in W_e$)
- $\langle \mathcal{A}_i(s), w \rangle \models_s \neg\varphi$ iff $\langle \mathcal{A}_i(s), w \rangle \not\models_s \varphi$
- $\langle \mathcal{A}_i(s), w \rangle \models_s \varphi \vee \psi$ iff $\langle \mathcal{A}_i(s), w \rangle \models_s \varphi$ or $\langle \mathcal{A}_i(s), w \rangle \models_s \psi$
- $\langle \mathcal{A}_i(s), w \rangle \models_s \varphi \wedge \psi$ iff $\langle \mathcal{A}_i(s), w \rangle \models_s \varphi$ and $\langle \mathcal{A}_i(s), w \rangle \models_s \psi$
- $\langle \mathcal{A}_i(s), w \rangle \models_s \varphi \Rightarrow \psi$ iff $\langle \mathcal{A}_i(s), w \rangle \not\models_s \varphi$ or $\langle \mathcal{A}_i(s), w \rangle \models_s \psi$
- $\langle \mathcal{A}_i(s), w \rangle \models_s \langle \mathcal{O}_i \rangle \varphi$ iff $\langle \mathcal{A}_i(s), w' \rangle \models_s \varphi$ for some w' such that $w \sim_{\mathbb{O}_s}^i w'$
- $\langle \mathcal{A}_i(s), w \rangle \models_s [\mathcal{O}_i] \varphi$ iff $\langle \mathcal{A}_i(s), w' \rangle \models_s \varphi$ for all w' such that $w \sim_{\mathbb{O}_s}^i w'$
- $\langle \mathcal{A}_i(s), w \rangle \models_s \mathbb{O}_p \varphi$ iff $\langle \mathcal{A}_i(s), w_p \rangle \models_s \varphi$ where w_p is the denotation of p .

Definition 8 (Observation accessibility relation) An expectation q is observable (reachable) from an expectation p , $p \sim_{\mathbb{O}_s}^i q$ iff there exists an observation method $o_k(\varepsilon_1, \dots, \varepsilon_k)$ where

- $\varepsilon_1, \dots, \varepsilon_k$ are valid at g_s
- $p \in \{\varepsilon_1, \dots, \varepsilon_k\}$
- $\exists \mathcal{E} \in o_k(\varepsilon_1, \dots, \varepsilon_k)$, such that $q \in \mathcal{E}$.

Hence $[\mathcal{O}_i] \varphi$ says from the current expectation, φ will hold after any available observation method is carried out. So if the eagle is currently expecting that it will see the sparrow, it will expect the proposition $q (= \varepsilon_4)$ – “the sparrow is caught” holds (i.e. observable) regardless of what observation methods “chasing the sparrow” or “staying in the cave” is taken. $\langle \mathcal{O}_i \rangle \varphi$ however says from the current expectation, there are only some observation methods that would bring φ into the agent’s mind.

The expectation operator \mathbb{O} is defined similarly to the observation operator \mathbb{O} . $\mathbb{O}_p \varphi$ hence asserts that there is an expectation p where φ holds. Hence, if p is the expectation “seeing a sparrow”, and r is “there is some light”, $\mathbb{O}_p r$ is read “There is some light whenever I expect to see a sparrow”.

$$\frac{\mathbb{O}_p \langle \mathcal{O}_i \rangle \varphi}{\mathbb{O}_p \langle \mathcal{O}_i \rangle a} (\diamond); \frac{\neg \mathbb{O}_p \langle \mathcal{O}_i \rangle \varphi; \mathbb{O}_p \langle \mathcal{O}_i \rangle q}{\neg \mathbb{O}_a \varphi} (\neg \diamond); \frac{\mathbb{O}_p [\mathcal{O}_i] \varphi \quad \mathbb{O}_p \langle \mathcal{O}_i \rangle q}{\mathbb{O}_q \varphi} (\square); \frac{\neg \mathbb{O}_p [\mathcal{O}_i] \varphi}{\mathbb{O}_p \langle \mathcal{O}_i \rangle a} (\neg \square)$$

Table 1. Some modality inference rules for observation system

The eagle’s deliberation can now be formalised using the rules in **Table 1**. Consider the case when the cave system consists of only one cave looping back on itself. The passageway is dark, but there is some light in the cave. The eagle

can only see the sparrow if there is some light $\mathcal{E}_p r$. When the sparrow flies into the dark passageway, the eagle would say I expect whatever I do, I can only see the sparrow in this cave s , $\mathcal{E}_p[\mathcal{O}_i]s$. Whenever I see the sparrow, I know a way to catch it $\mathcal{E}_p\langle\mathcal{O}_i\rangle q$. Using the (\Box) rule the eagle will decide to stay in s to catch the sparrow $\mathcal{E}_q s$.

However, if the cave system consists of an infinite number of caves linking by unidirectional passageways, the deliberation will be slightly changed. After some observations, the eagle would discover the next cave cannot be s ($\neg @_s\langle\mathcal{E}_i\rangle s$), but another cave t ($@_s\langle\mathcal{E}_i\rangle t$). Hence, when the sparrow disappears, the eagle would say, I expect whatever I do, I can only observe the sparrow in the next cave $\mathcal{E}_p[\mathcal{O}_i]\langle\mathcal{E}_i\rangle t$. Whenever I see the sparrow, I know a way to catch it $\mathcal{E}_p\langle\mathcal{O}_i\rangle q$. Using the (\Box) rule, we will be able to derive $\mathcal{E}_q\langle\mathcal{E}_i\rangle t$, which suggests to capture the sparrow, the eagle should follow the sparrow into the passageway linking to the next cave t .

4 Labelled BDI logics

Rao and Georgeff's *Belief-Desire-Intention* (BDI) logics are one of the most successful theories for agent specification or verification languages in agent research community. Following the philosopher Bratman [7], a formalisation of the three mental attitudes *belief*, *desire*, *intention* and their interactions has been investigated as characterisations of an agent. Most work on BDI logics focused on possible relationships between these three mental attitudes by adding different constraints on their interactions. According to Bratman [7], assuming an eagle is a rational agent, the eagle will not intend to catch the sparrow if it believes the sparrow is uncatchable. But it still tries its best (intends) to catch the sparrow for hunger though it does not believe it can catch it (*asymmetry thesis*). Also, the eagle may believe it can catch the sparrow, but it is not necessary that it intends to catch a sparrow now (*non-transference principle*). Additionally, if the eagle intends to catch a sparrow for hunger, though it believes catching the sparrow is certainly energy burning, it will not intend to burn out its energy (*side-effect free principle*). Rao and Georgeff [17] formally put these constraints in the following proposition:

Proposition 1 *A rational agent a_i must satisfy the following principles:*

- *Asymmetry thesis: An agent cannot have beliefs inconsistent with intentions, but can have incomplete beliefs about its intentions.*
 - (BI-ICN) $\not\models [\mathcal{I}_i]\varphi \wedge [\mathcal{B}_i]\neg\varphi$
 - (BI-ICM) $\exists \mathfrak{M}, \mathfrak{M} \models [\mathcal{I}_i]\varphi \wedge \neg[\mathcal{B}_i]\varphi$
- *Non-transference principle: An agent who believes φ should not be forced to intend φ .*
 - (BI-NT) $\exists \mathfrak{M}, \mathfrak{M} \models [\mathcal{B}_i]\varphi \wedge \neg[\mathcal{I}_i]\varphi$
- *Side-effect free principle: if an agent intends φ and believes that $\varphi \Rightarrow \psi$, it should not be forced to intend the side-effect ψ .*
 - (BI-SE) $\exists \mathfrak{M}, \mathfrak{M} \models [\mathcal{I}_i]\varphi \wedge [\mathcal{B}_i](\varphi \Rightarrow \psi) \wedge \neg[\mathcal{I}_i]\psi$

These constraints also apply for belief-goal, and goal-intention pairs.

The constraints between these mental attitudes are set based on the relationships between the accessibility relations. Three well known cases of these systems were studied by Cohen and Levesque in term of *realism* ($\mathbb{B}_i \subseteq \mathbb{G}_i$) [8], by Rao and Georgeff in terms of *strong realism* ($\mathbb{G}_i \subseteq \mathbb{B}_i$) [18] and *weak realism* ($\mathbb{G}_i \cap \mathbb{B}_i \neq \emptyset$) [17]. Rao and Georgeff [17] also concluded that weak realism is the only system that satisfies all desirable properties of a rational agent.

The real world model of BDI logics is viewed as a single past-branching time future tree. Each possible world of any belief, desire (goal) or intention models consists of a subtree of the above temporal structure. In other words, they are different images of the real world in the agent's mind. However, a major drawback of BDI logics is that it is very unclear which part of the real world temporal structure should be in an agent's mental attitudes, beliefs, desires, or intentions. There is no formal correspondence from the mental models to the world structure. Consider the eagle chasing the sparrow again. BDI language is unable to tell when a particular event would happen. Thus the sentence "eventually the sparrow will be caught" ($[\mathcal{B}_i]\diamond q$) can be interpreted to be true at two different time points t_1 and t_2 . Regardless of how many observations it can take, the eagle using BDI logics is unable to tell if t_1 and t_2 is a unique time point. A BDI agent would continue to seek for a sparrow after having one caught. Expectation-observation logic however allows us to tell if these points are equal by $@_{t_1}t_2$. Hence, if it is the case, the eagle will drop all subsequent goals to catch the sparrow in the future in the cave system.

A translation from BDI languages to our expectation-observation language can be useful to attain the new expressive power. Firstly, we can construct our observation system using similar temporal structure. Our expectation modality in the system becomes the expectation about the future, equivalently to future modality.

Definition 9 (Mental translation) *A mental translation taking BDI formulae to expectation-observation formulae is defined as follows:*

Expectation model	Attention model
$-\Box\varphi \stackrel{def}{=} [\mathcal{E}_i]\varphi$	
$-\Diamond\varphi \stackrel{def}{=} \langle \mathcal{E}_i \rangle \varphi$	
$-\varphi \mathcal{U} \psi \stackrel{def}{=} \langle \mathcal{E}_i \rangle (s \wedge \psi) \wedge [\mathcal{E}_i] (\langle \mathcal{E}_i \rangle s \Rightarrow \varphi)$	$-\ [\mathcal{B}_i]\varphi \stackrel{def}{=} [\mathcal{O}_i^{\mathcal{B}}]\varphi$
$-\ \bigcirc\varphi \stackrel{def}{=} \langle \mathcal{E}_i \rangle (s \wedge \varphi) \wedge [\mathcal{E}_i] (\neg \langle \mathcal{E}_i \rangle s)$	$-\ [\mathcal{G}_i]\varphi \stackrel{def}{=} [\mathcal{O}_i^{\mathcal{G}}]\varphi \wedge \neg\varphi$
	$-\ [\mathcal{I}_i]\varphi \stackrel{def}{=} \langle \mathcal{O}_i^{\mathcal{I}} \rangle p \wedge @_p\varphi$

where $W_e^{\mathcal{B}} = \{p \in \Phi \mid @_s p\}$, $W_e^{\mathcal{G}} = W_e^{\mathcal{I}} = \{p \in \Phi \mid @_s \langle \mathcal{E}_i \rangle p\}$

The crucial difference between $\mathcal{O}_i^{\mathcal{B}}$, $\mathcal{O}_i^{\mathcal{G}}$, $\mathcal{O}_i^{\mathcal{I}}$ is only based on which primitive expectations are selected into the sets of possible worlds $W_e^{\mathcal{B}}$, $W_e^{\mathcal{G}}$, $W_e^{\mathcal{I}}$. At a particular world named as s , beliefs are its mental states, where goals and intentions are its mental states about what it expects to happen next if it takes more observation.

By this translation, *beliefs* now are an agent's expectations of what will be observable. If there is no observation linking to the expectation, a belief may well

be false. *Goals* are what an agent expects to be observable (in future), but are not observable now. This definition satisfies a number of required properties for goals [20]. Observable also means ‘achievable’ or ‘possible’ – the agent only has goals that it believes achievable. However, ‘ $\neg\varphi$ ’ guarantees the goal is *unachieved* or at least expected to be unachieved. Consistency and persistence are just normal logical properties of goals. Although *intention* could be defined as $\langle \mathcal{O}_i^T \rangle \varphi$, the above definition insists the agent has committed to a specific observation method o_k to achieve φ at the subsequent expectation p .

The definition also satisfies weak-realism constraint. The overlapping between beliefs and goals is the set of expectations that hold now and at some time points in the future ($@_s p$ and $@_s \langle \mathcal{E}_i \rangle p$). However, the agent does not have direct observation of the expectations at the current observation. The non-overlapped part of goals are what the agent does not expect to observe now ($@_s \neg p$), but it expects them to be observable in future ($@_s \langle \mathcal{E}_i \rangle p$). Conversely, a belief is not in the goal set if there is a direct observation now ($@_s p$) or the agent expects it will not happen at all ($\neg @_s \langle \mathcal{E}_i \rangle p$).

Similarly, the overlapping between beliefs and intentions is where the subsequent expectation p is also in the set of expectations of the current observation. An intentions will no longer be in the set of beliefs if the resultant expectation of the observation method o_k which is committed to the intention, does not hold at the current observation g_s . On the other hand, a belief is out of the intention set if there is no observation method o_k links to p . In other words, the agent believes φ is observable, but it has no way to observe φ now.

The overlapping between goals and intentions is where the expectation p is in the set of expectations of all possible observations about to occur in the whole system. An intention may not be a goal if it is already observable at the current observation. On the other hand, a goal will not be a specific intention if the agent expects φ can only be achieved by other observation methods not the one associated with the intention.

By this definition, it is clear that the following proposition holds:

Proposition 2 *The agent modelled by the above observation-expectation system is a rational agent. That is, it satisfies asymmetry thesis, side-effect free and non-transference principles.*

Proof. See **Appendix A**.

5 Conclusion and further work

In this paper we apply hybrid logic to address a well-known unresolved problem in the agent research community, the computational grounding problem and to introduce a formal correspondence between mental and computational models. It is certainly not yet another paper about hybrid logics. Hence, we do not show decidability, completeness results which have been deeply studied by other researchers [6], [1], [3], [4]. Instead, our crucial argument here is that any concrete computational models are extensional whereas any mental models are intensional. Agent specifications using orthodox modal languages can only express intensional aspects and therefore fail to make connection to extensional aspects.

Apart from BDI logics discussed above, a major strand of research led by the work of Fagin et al. [11], [10] has attempted to bring the external to the internal perspective using interpreted systems as the basis of epistemic logic. This approach tightly connects the internal to the external. The approach hence started from a perfectly synchronised mental model with the environment where everything is directly reflected to every agent’s mind. Then, the connection is loosened to reflect the fact that agents are imperfect. A dilemma has arisen in this investigation [10, Chapter 11], simultaneity (time synchronisation) strongly affects the attainability of (common) knowledge, but true simultaneity cannot be attained in reality. Interestingly, the resolution of this paradox leads to the ability to record time points and the granularity of time – timestamped (common) knowledge. However, unlike hybrid languages, their naming mechanisms cannot be manipulated and hence reasoned as formulae. An extension for time observation in our framework to link with this work hence appears very promising.

An extension of Fagin et al.’s interpreted systems, \mathcal{VSK} systems and \mathcal{VSK} logic [23], provided another attempt to formalise the connection between the states of agents within a system and the percepts received by them. The imperfect situation is captured by using the notion of “partial observability” in POMDPs [15] through \mathcal{V} and \mathcal{S} modalities. Their knowledge modality \mathcal{K} remains the same as modal epistemic logic [10]. There are two crucial drawbacks of this work. Firstly, *visibility* function is similar to *observation function* by van der Meyden [19] which is only capable of capturing the discrete states of an environment but not the relationships between them. Secondly, \mathcal{VSK} fails to fully capture human perception which can be faulty. Our framework using the idea from hybrid logic overcomes the former problem by letting an observation about a relationship be mapped onto an expectation of named state (e.g. $@_s\langle\mathcal{E}_i\rangle t$). The second problem is resolved by adding hypothetical observations from agents’ effectors into the concept of observability.

Finally, our chief further work is to show completeness and correspondence results. However it is also worth noting that our work is principally based upon fibring techniques by Gabbay [12] and analytic deduction via labelled deductive systems **LKE** by D’Agostino and Gabbay [9]. These works provide a potential approach towards a uniform way of combining logical systems, hence modelling cooperative reasoning in interactive dynamic environment.

References

1. C. Areces, P. Blackburn, and M. Marx. Hybrid logics: Characterization, interpolation and complexity. *Journal of Symbolic Logic*, 66(3):977 – 1010, 2001.
2. P. Blackburn. Nominal tense logic. *Notre Dame Journal of Formal Logic*, 34(1):56–83, 1993.
3. P. Blackburn. Internalizing labelled deduction. *Journal of Logic and Computation*, 10:137–168, 2000.
4. P. Blackburn. Representation, reasoning, and relational structures: a hybrid logic manifesto. *Logic Journal of the IGPL*, 8(3):339–625, 2000.

5. P. Blackburn, M. de Rijke, and Y. Venema. *Modal logic*. Cambridge University Press, 2001.
6. P. Blackburn and M. Tzakova. Hybrid languages and temporal logic. *Logic Journal of the IGPL*, 7:27–54, 1999.
7. M. E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, 1987.
8. P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.
9. M. D’Agostino and D. Gabbay. A generalization of analytic deduction via labelled deductive systems. Part I: Basic substructural logics. *Journal of Automated Reasoning*, 13:243–281, 1994.
10. R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. The MIT Press, Cambridge, Massachusetts, 1995.
11. R. Fagin, J. Y. Halpern, and M. Y. Vardi. What can machines know? on the properties of knowledge in distributed systems. *Journal of the ACM*, 39(2):328–376, 1992.
12. D. Gabbay. *Fibring Logics*, volume 38 of *Oxford Logic Guides*. Oxford University Press, 1999.
13. P. B. Gove, editor. *Webster’s Revised Unabridged Dictionary*. Merriam Webster Inc., 3rd edition, 2002.
14. J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of The Two Notions*. Cornell University Press, Ithaca, New York, 1962.
15. L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998.
16. A. Rao. AgentSpeak(L): BDI Agents speak out in a logical computable language. In W. Van de Velde and J. Perram, editors, *Proceedings of the Seventh Workshop on Modelling Autonomous Agents in Multi-Agent World (MAAMAW’96)*, volume 1038 of *Lecture Notes in Artificial Intelligence*, pages 42–55, Eindhoven, The Netherlands, 1996. Springer-Verlag.
17. A. Rao and M. Georgeff. Asymmetry thesis and side-effect problems in linear-time and branching-time intention logics. In J. Myopoulos and R. Reiter, editors, *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 498–505, Sydney, Australia, 1991. Morgan Kaufmann publishers Inc.: San Mateo, CA, USA.
18. A. Rao and M. Georgeff. Modelling rational agents within a BDI-architecture. In R. Fikes and E. Sandewall, editors, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, pages 473–484, Cambridge (USA), 1991.
19. R. van der Meyden. Common knowledge and update in finite environments. *Information and Computation*, 140(2):115–157, 1998.
20. M. Winikoff, L. Padgham, J. Harland, and J. Thangarajah. Declarative and procedural goals in intelligent agent systems. In *Eighth International Conference on Principles of Knowledge Representation and Reasoning*, pages 470–481, 2002.
21. M. Woolridge. Computationally Grounded Theories of Agency. In E. H. Durfee, editor, *Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS 2000)*, volume 9. IEEE Press, 2000.
22. M. Woolridge. *An introduction to MultiAgent System*. John Wiley & Sons, Chichester, England, 2002.
23. M. Woolridge and A. Lomuscio. Reasoning about visibility, perception, and knowledge. In N. Jennings and Y. Lespérance, editors, *Intelligent Agents VI*, volume Lecture Notes in AI Volume. Springer-Verlag, 2000.

A Proofs

	BDI language	Expectation-observation language
BI-ICN	$\not\models [I_i]\varphi \wedge [B_i]\neg\varphi$	$\not\models \langle \mathcal{O}_i^I \rangle p \wedge \mathcal{E}_p \varphi \wedge [\mathcal{O}_i^B]\neg\varphi$
BI-ICM	$\exists \mathfrak{M}, \mathfrak{M} \models [I_i]\varphi \wedge \neg[B_i]\varphi$	$\exists \mathfrak{M}, \mathfrak{M} \models \langle \mathcal{O}_i^I \rangle p \wedge \mathcal{E}_p \varphi \wedge \langle \mathcal{O}_i^B \rangle q \wedge \mathcal{E}_q \neg\varphi$
GI-ICN	$\not\models [I_i]\varphi \wedge [G_i]\neg\varphi$	$\not\models \langle \mathcal{O}_i^I \rangle p \wedge \mathcal{E}_p \varphi \wedge [\mathcal{O}_i^G]\neg\varphi \wedge \varphi$
GI-ICM	$\exists \mathfrak{M}, \mathfrak{M} \models [I_i]\varphi \wedge \neg[G_i]\varphi$	$\exists \mathfrak{M}, \mathfrak{M} \models \langle \mathcal{O}_i^I \rangle p \wedge \mathcal{E}_p \varphi \wedge \langle \mathcal{O}_i^G \rangle q \wedge (\mathcal{E}_q \neg\varphi \vee \varphi)$
BG-ICN	$\not\models [G_i]\varphi \wedge [B_i]\neg\varphi$	$\not\models ([\mathcal{O}_i^G]\varphi \wedge \neg\varphi) \wedge [\mathcal{O}_i^B]\neg\varphi$
BG-ICM	$\exists \mathfrak{M}, \mathfrak{M} \models [G_i]\varphi \wedge \neg[B_i]\varphi$	$\exists \mathfrak{M}, \mathfrak{M} \models ([\mathcal{O}_i^G]\varphi \wedge \neg\varphi) \wedge \neg[\mathcal{O}_i^B]\varphi$
BG-NT	$\exists \mathfrak{M}, \mathfrak{M} \models [B_i]\varphi \wedge \neg[G_i]\varphi$	$\exists \mathfrak{M}, \mathfrak{M} \models [\mathcal{O}_i^B]\varphi \wedge (\neg[\mathcal{O}_i^G]\varphi \vee \varphi)$
BI-NT	$\exists \mathfrak{M}, \mathfrak{M} \models [B_i]\varphi \wedge \neg[I_i]\varphi$	$\exists \mathfrak{M}, \mathfrak{M} \models [\mathcal{O}_i^B]\mathcal{B} \wedge ([\mathcal{O}_i^I]\neg p \vee \neg\mathcal{E}_p \varphi)$
GI-NT	$\exists \mathfrak{M}, \mathfrak{M} \models [G_i]\varphi \wedge \neg[I_i]\varphi$	$\exists \mathfrak{M}, \mathfrak{M} \models ([\mathcal{O}_i^G]\varphi \wedge \neg\varphi) \wedge ([\mathcal{O}_i^I]\neg p \vee \neg\mathcal{E}_p \varphi)$
BI-SE	$\exists \mathfrak{M}, \mathfrak{M} \models [I_i]\varphi \wedge [B_i](\varphi \Rightarrow \psi) \wedge \neg[I_i]\psi$	$\exists \mathfrak{M}, \mathfrak{M} \models (\langle \mathcal{O}_i^I \rangle p \wedge \mathcal{E}_p \varphi \wedge \langle \mathcal{O}_i^B \rangle q \wedge \mathcal{E}_q \neg\varphi \wedge \mathcal{E}_p \neg\psi) \vee (\langle \mathcal{O}_i^I \rangle p \wedge \mathcal{E}_p \varphi \wedge [\mathcal{O}_i^B]\psi \wedge \mathcal{E}_p \neg\psi)$
GI-SE	$\exists \mathfrak{M}, \mathfrak{M} \models [I_i]\varphi \wedge [G_i](\varphi \Rightarrow \psi) \wedge \neg[I_i]\psi$	$\exists \mathfrak{M}, \mathfrak{M} \models (\langle \mathcal{O}_i^I \rangle p \wedge \mathcal{E}_p \varphi \wedge \langle \mathcal{O}_i^G \rangle q \wedge \mathcal{E}_q \neg\varphi \wedge \mathcal{E}_p \neg\psi) \vee (\langle \mathcal{O}_i^I \rangle p \wedge \mathcal{E}_p \varphi \wedge \varphi \wedge \mathcal{E}_p \neg\psi) \vee (\langle \mathcal{O}_i^I \rangle p \wedge \mathcal{E}_p \varphi \wedge [\mathcal{O}_i^G]\psi \wedge \mathcal{E}_p \neg\psi)$
BG-SE	$\exists \mathfrak{M}, \mathfrak{M} \models [G_i]\varphi \wedge [B_i](\varphi \Rightarrow \psi) \wedge \neg[G_i]\psi$	$\exists \mathfrak{M}, \mathfrak{M} \models ([\mathcal{O}_i^G]\varphi \wedge \neg\varphi \wedge \neg[\mathcal{O}_i^B]\varphi \wedge \neg[\mathcal{O}_i^G]\psi \vee \psi) \vee ([\mathcal{O}_i^G]\varphi \wedge \neg\varphi \wedge [\mathcal{O}_i^B]\psi \wedge (\neg[\mathcal{O}_i^G]\psi \vee \psi))$

Table 2. BDI constraints in expectation observation language

From **Table 2**, it is clear that (BI-ICN), (GI-ICN) and (BI-ICN) constraints are satisfied by our framework. (BI-ICM) happens when intention is not in the belief set. From the translation, it says p, q are indistinguishable to an observation method and they reside separately in the two sets intention and belief respectively. Similarly for goal-intention pair except that the agent can intend an achieved goal (i.e. no longer goal) for example to maintain its achievement. (BG-ICM) may look counter-intuitive. However, our translation insists the difference. Beliefs are based on the current observations only, where goals can come from different sources (from other agents – e.g. your boss). Hence, this constraint seems appropriate in a multi-agent system. The asymmetry thesis principle is preserved under the new language.

(BG-NT) appears obvious, since φ is observable now, the agent can hold a belief about φ without having φ as its goal. The emphasis of the commitment to an intention can now be used for (BI-NT) and (GI-NT). Commitment ties a specific mental state p to an intention. Therefore, the agent will not intend φ but it can still believe or have φ as goal. For example, a person does not intend war in Iraq, but believes war is there. This also seems intuitive in a multi-agent environment. The non-transference principle is hence preserved under expectation observation logic.

(BI-SE) can be rewritten as $([I_i]\varphi \wedge \neg[B_i]\varphi \wedge \neg[I_i]\psi) \vee ([I_i]\varphi \wedge [B_i]\psi \wedge \neg[I_i]\psi)$ which appears to be a restricted version of (BI-ICM) and (BI-NT). So if the agent's belief is incomplete about the intention and at the intended expectation p , ψ does not hold, the agent would not worry about the side effect. On the other hand, assuming the agent believes ψ , according to (BI-NT) it is not forced to intend ψ . The translation clarifies the situations where side-effect free can be satisfied. We can achieve similar results for (GI-SE) and (BG-SE).