Expectation: The Logic of Flexible Motivation

Binh Vũ Trần	James Harland	Margaret Hamilton
tvubinh@cs.rmit.edu.au	jah@cs.rmit.edu.au
mh@cs.rmit.edu.au
School of Computer Science and Information Technology
RMIT University, Australia

ABSTRACT
The notion of goals, which tend to have strong properties (e.g. goals should be consistent, feasible, not yet achieved, etc.) has received more attention than other motivational notions such as desires. Desires seem more applicable in a multi-agent setting since an agent may cooperate with other agents without an explicitly agreed goal, but cannot act purely independently. However, the connection between desires and real world actions is not formally described. In this work, we introduce the notion of expectation in order to address these problems. A formalism is developed to describe the association of expectations and observations as well as how expectations could drive future observations.

Categories and Subject Descriptors
I.2.0 [Artificial Intelligence]: General Philosophical foundations; I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods—Modal logic, Temporal logic; I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence—Intelligent agents, Multiagent systems

General Terms
Theory

Keywords
Expectation, Observation, Logical Models of Observation and Expectations

1. INTRODUCTION
The mechanisms to generate actions appropriately for co-ordination in a multi-agent environment are usually formalised through the motivational attitudes of agents. Formalisms for goals are often strong with properties such as unachieved, possible, persistent, consistent etc. which may no longer hold due to the dynamics and unpredictability of an environment. A notion with weaker properties such as desire seems more applicable. Unfortunately, desires possess few logical properties [2]. More importantly, modal languages, the agent specifications, have a major drawback – they allow agents to reason about their intensional aspects but not about the extensional aspects. Hence, this hampers us from formally describing the connection between the intensional – motivational attitudes – and the extensional – computational models.

Imagine an eagle is chasing a sparrow through a cave system where every cave appears identical and has 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 whether the cave system has an infinite number of caves connected with each other 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 stay still in the current cave waiting for the sparrow in the latter case.

In this paper, we believe a combination of expectation and observation is essential for modelling agents in a multi-agent environment. Hence, following the idea of internalising the extensional in hybrid logics [1], we develop a formalism of expectations as being the mental states that are associated with observations bridging the real world and mind. Although expectations are derived from observations, based on inferential mechanism, expectations can also drive an agent’s future observations. This allows the modelling of the flexible change of an agent’s motivations according to which observations it takes.

2. OBSERVATION-EXPECTATION SYSTEM
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 \((g,\varepsilon)\) which links a global state \(g \in G\) with a mental state \(\varepsilon \in E\). Unfortunately, this only connection to the real world is not always available for 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 attention model accordingly. Thus, when chasing the sparrow, if the sparrow were to disappear into a passageway, the eagle could 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.

In any observation system \(OS\), sensors \(S\) and effectors \(E\) 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 $O_0 = S \cup E$. A more complicated set of observation methods $O_k$ 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.

**Definition 1. (Observation methods)** An observation method family is a set of observation method sets $\mathcal{O} = \{O_k\}_{k \in \mathbb{N}}$ where $O_k$ is a set of observation methods of arity $k$ for every $k \in \mathbb{N}^+$. $O_k$ is inductively defined as follows:

- $\varepsilon \in \mathcal{E}$ for all $\varepsilon \in o_0, \forall o_k \in O_k$
- $o_k(\varepsilon_1, \ldots, \varepsilon_k) \subseteq \varepsilon(\varepsilon), \forall o_k \in O_k$ and $\varepsilon_1, \ldots, \varepsilon_k \in \mathcal{E}$

Intuitively, whilst adopting any observation method $o_k$, an agent may have various observations about a real world state $g$ which bring various expectations $\{\mathcal{E}, \mathcal{E}', \mathcal{E}''\}$ 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 $\mathcal{E}, \mathcal{E}'$ will appear indistinguishable to an individual agent if its expectations in both sets do not change.

Hence, for any agent $a_i$ identified by $i \in I$, given a primitive expectation proposition $F = \{p, q, r, \ldots\}$, an observation naming set $\Xi = \{s, t, \ldots\}$, the observation interpretation function $\pi : (\Phi \cup \Xi) \rightarrow \varphi(s)$ based on available set of observations tells which expectation is associated with which global state. Following hybrid logic [1], the major distinction is that $\pi$ returns a singleton for every $s \in \Xi$. Hence, $s$ is true at a unique global state, and tags this state. An expectation model $M = (OS, \pi)$ describes how expectations are generated from observations. The images $g, g'$ of the world are linked by an expectation accessibility relation $g \sim_{\pi} g'$ if $\exists g, \exists g' \subseteq g, \exists g' \subseteq g$, such that $\mathcal{E}, \mathcal{E}'$ are indistinguishable to the agent $a_i$ through an arbitrary existing observation method $o_k(\varepsilon_1, \ldots, \varepsilon_k)$, where $\varepsilon_1, \ldots, \varepsilon_k \in g$ and $\varepsilon_1, \ldots, \varepsilon_k \in g'$.

An attention model $A(s) = (W_e, \sim_{\pi}, \rho_s)$ on the other hand, based on a set of primitive expectations $W_e$ at a global state $s$, describes the attentional observations of the agent. The function $\rho_s : \Xi \rightarrow \varphi(W_e)$ interprets what are possible expectations for an observation at $s$. An expectation $q$ is observable (reachable) from an observation $p$, if there exists an observation method $o_k(\varepsilon_1, \ldots, \varepsilon_k)$ such that $\varepsilon_1, \ldots, \varepsilon_k$ are valid at $g$, $p \in \{\xi_1, \ldots, \xi_k\}$, and $\exists \xi \in o_k(\xi_1, \ldots, \xi_k)$, such that $\xi, \xi \in \mathcal{E}$.

**Definition 2.** Other than the standard part, the semantics of observation-expectation logic are defined as follows:

1. $[M], (g) = (s), \varphi$ if $f \varphi \bar{g}$, where $g \sim_{\pi} g'$, $[M], (g') = \varphi$
2. $[M], (g) = (s), \varphi[\mathcal{M}], (g') = \varphi \forall g', g \sim_{\pi} g'$
3. $[M], (g) = s \forall s \in \Xi, \pi(s) = \{g\}$
4. $[M], (g) = \exists\phi \varphi[\mathcal{M}], (g_s) = \varphi, g_s$ is the denotation of $s$.
5. $[A(s), w] = \{t \mid \forall w \in \rho_s(t), \forall t \in \Xi\}$
6. $[A(s), w] = \{p \mid \forall p \in \rho_s(p) (\forall w \in W_e)\}
7. [A(s), w] = \{\mathcal{O}(s), \varphi[\mathcal{O}(s), w') = \varphi, \forall w', w \sim_{\pi} w'\}$
8. $[A(s), w) = \{\mathcal{O}(s), w') = \varphi, \forall w', w \sim_{\pi} w'\}$
9. $[A(s), w] = \ast \mathcal{O}(s), \varphi[\mathcal{O}(s), w] = \varphi$.

Thus, if $[E_s], \varphi$ is true in some state $g \in G$, then by adopting observation method $o_k$, the local states of the agent $a_i$ about the environment (its expectations) remains the same.

**3. DISCUSSION AND FURTHER WORK**

A key idea to differentiate our work from BDI logic [4] and epistemic logic [3] is that the expectation accessibility relation and observation interpretation function are purely dependent on the observations that an agent has actually made. By formally describing the connection between expectation and observation, we have attempted to close the BDI gap and offered an ability to describe faulty human perception in epistemic logic. Our current chief investigation is to explore a uniform way of combining expectation and observations in a dynamic, unpredictable environment.

**4. REFERENCES**


