

Technical Report IIIA 2012-03

January 2012

***An Experience-Based BDI Logic:
Motivating Shared Experiences and
Intentionality***

Nardine Osman
Mark d'Inverno
Carles Sierra



Campus de la Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona



An Experience-Based BDI Logic: Motivating Shared Experiences and Intentionality

Nardine Osman¹ and Mark d’Inverno² and Carles Sierra¹

¹ Artificial Intelligence Research Institute (IIIA-CSIC), Barcelona, Spain

² Computing Department, Goldsmiths, University of London

Abstract

This paper introduces the notion of experiences, which help situate agents in their environment, providing a concrete link on how the continually evolving environment impacts the evolution of an agent’s BDI model. Then, using the notion of shared experience as a primitive construct, we develop a novel formal model of shared intention which we believe more adequately describes and motivates social behaviour than traditional BDI logics that focus on modelling individual agents. Whilst many philosophers have strongly argued that collective intentionality cannot always be equated to the collection of the individual agents, there has been no AI model that has proposed how this could occur. To the best of our knowledge this is the first attempt to develop the notion of shared experience from an AI perspective that cannot be reduced to descriptions of a single agent.

Introduction

Existing models of BDI agents typically assume the agent has its belief base already built (d’Inverno et al., 2004; Bordini, Hübner, and Wooldridge, 2007; Winikoff, 2005). The abstract logics describing BDI systems do not consider how an agent’s ongoing activity in an environment builds these elements and, moreover, how that activity impacts on the relationships between them. In this work, we provide an agent model that takes into consideration how the environment and activities in the environment helps shape the belief base of an agent situated in that environment. This includes perceiving environments that results in agents having new experiences, which in turn affects the perceiving agent’s beliefs, desires, and intentions, which then leads to new actions, and back round again to new experiences.

After presenting the basic agent model, we introduce a logical formalism (X-BDI) to describe agent reasoning whose semantics are grounded in the model. In this formalism we define the notion of shared experience based upon the notion of individual experience, and use it to address the issue of collective intentionality. We believe shared experiences and shared intentions provide the basis of social behaviour which is the foundation of multiagent system research. Our model provides an adequate description of why

agents would approach others to interact with in the first place outside the concept of a shared plan.

The remainder of this paper is organised as follows: we start with a specification of the agent model (which provides the grounding for the logic’s computational semantics), then we introduce the logic for the experience based BDI model (X-BDI), and follow it with an example that illustrates the impact of this work. We end with a brief literature review and some concluding remarks.

The Agent Model

We argue that the physical world is populated with objects (some of which may be labelled as agents) and changed through events (where some subset may be labelled as actions performed by agents). Here, our focus is not on the physical world itself, but on how it is perceived by agents. It is the perception of the agent situated in a given environment that affects the agent’s beliefs, desires, and intentions. More precisely, we introduce the notion of experience, which is the result of the agent perceiving its environment and forming some conceptions about it. We argue that experiences shape beliefs (usually resulting in belief updates), beliefs influence desires, and desires drive intentions. Intentions in turn lead the agent to act in its environment (where the action could be as simple as to observe), which results in the agent having new experiences. In other words, we say that the experience is the missing link that completes the BDI cycle by situating the agent in its environment and illustrating how the environment impacts the evolution of an agent’s BDI. This section provides the basic definitions of experiences, beliefs, desires, and intentions. We note that our proposal is based on the assumption that an agent can perceive its environment, resulting in having a set of percepts. Although we note that percepts may sometimes be imagined (following the cognitive neuroscience view that considers imagination as a second order perception (Schütz-Bosbach and Prinz, 2007; Perrett et al., 2009; Weiler, Suchan, and Daum, 2010; Szpunar, Watson, and McDermott, 2007)). As such, the set of agent a ’s percepts is referred to as $\mathcal{P}_a = \mathcal{R}_a \cup \mathcal{I}_a$, where percepts are either real (\mathcal{R}) or imagined (\mathcal{I}). We also note that percepts, both real and imagined, could be described various types of elements, amongst these are percepts of agents (\mathcal{AG}) and percepts of actions (\mathcal{AC}).

Experiences

An experience is defined by a number of percepts that the agent decides to group together and label as a single entity. Experiences may be basic (composed of percepts) or compound (composed of a combination of percepts and other experiences). An example of the former would be the sunset one witnessed in Phuket. An example of the latter would be going out last night, which contains nested experiences such as having a drink, seeing a show, and having dinner.

We adopt Kant’s argument (Kitcher, 1990) that the basic cognitive hard wired relationships are time ordering (\prec), causality (\rightsquigarrow), and spatial (\bowtie) relationships. We also adopt Kant’s view (Kant, 1838) that human beings follow the categorical imperative method that makes them classify things as good or bad. As such, we introduce the evaluation function (\heartsuit) for evaluating experiences. We note that one may also think of evaluations to be used to describe one’s emotions towards an experience. For instance, “I loved that game”.

We state that the set of all experiences of agent a is \mathcal{E}_a . A single experience of agent a , $E_a \in \mathcal{E}_a$, is then defined as the tuple:

$$E_a = \langle L, \mathcal{E}'_a \cup \mathcal{P}'_a, \prec, \rightsquigarrow, \bowtie, \heartsuit \rangle$$

- L describes the experience’s unique label that identifies it.
- $\mathcal{E}'_a \cup \mathcal{P}'_a$ describes its content (referenced as $\text{content}(E_a)$), which may be composed of percepts ($\mathcal{P}'_a \subseteq \mathcal{P}_a$), sub-experiences ($\mathcal{E}'_a \subseteq \mathcal{E}_a$), or a combination of both.
- $\prec \subseteq (\mathcal{E}'_a \cup \mathcal{P}'_a) \times (\mathcal{E}'_a \cup \mathcal{P}'_a)$ defines a partial temporal order over the sub-experiences and perceptions.
- $\rightsquigarrow \subseteq (\mathcal{E}'_a \cup \mathcal{P}'_a) \times (\mathcal{E}'_a \cup \mathcal{P}'_a)$ defines a causal relationship over the sub-experiences and perceptions, and it should satisfy the property $\rightsquigarrow \subseteq \prec$.
- $\bowtie \subseteq (\mathcal{E}'_a \cup \mathcal{P}'_a) \times (\mathcal{E}'_a \cup \mathcal{P}'_a) \times \mathcal{S}$ defines a spacial relationship between the sub-experiences and perceptions, and it should satisfy the property $(a, b, s) \in \bowtie \Rightarrow (a, b) \notin \prec \wedge (b, a) \notin \prec$. An example of the set of spacial relationships could be $\mathcal{S} = \{\text{above, below, left, right, behind, in front}\}$.
- $\heartsuit : \mathcal{L} \rightarrow \mathcal{V}$ defines the experience’s evaluation, which maps a label to an evaluation space. An example of an evaluation space is $\mathcal{V} = \{\text{positive, negative}\}$. The label is intended to represent the evaluation criteria, such as $\mathcal{L} = \{\text{impact, usefulness, contentment, surprise}\}$. The evaluation criteria may either represent rational criteria, such as whether “the experience helps one achieve their goals”, or emotional ones. One may think of numerous emotional criteria, but the six basic emotions according to Paul Ekman (Ekman, 1972) are: anger, disgust, fear, surprise, sadness, and happiness.

We distinguish a special type of experience, a basic experience \bar{E}_a , in which all the contents of the experience are single percepts: $\bar{E}_a = \langle L, \mathcal{P}'_a, \prec, \rightsquigarrow, \bowtie, \heartsuit \rangle$. Agent a ’s set of all basic experiences is referred to as $\bar{\mathcal{E}}_a$.

Another interesting type of basic experience is the basic action, which we define as a basic experience, whose content contains only the percepts describing the pre-conditions of the action, the action percept describing the action itself

($AC \in \mathcal{AC}$), and the percepts describing the post-conditions of the action. The temporal relation then states that all pre-condition percepts should precede the action percept and that the action percept should precede all the post-condition percepts. Additionally, the causal relation states that the action percept leads to the post-condition percepts. A basic action is then defined as a basic experience \bar{E}_a that satisfies the following properties:

$$\forall x \in \text{content}(\bar{E}_a) \cdot (\exists y \in \text{content}(\bar{E}_a) \cdot x \prec y \vee y \prec x)$$

$$x \prec y \wedge y \prec z \Rightarrow y \in \mathcal{AC}_a$$

$$|\mathcal{AC}_a \cap \text{content}(\bar{E}_a)| = 1$$

$$x \rightsquigarrow y \Rightarrow x \in \mathcal{AC}_a \wedge y \notin \mathcal{AC}_a$$

$$x \rightsquigarrow y \Rightarrow x \prec y$$

Beliefs

An agent holds beliefs about itself, its environment, its feelings (if any), and so on. There are different types of beliefs, such as: *factual beliefs* which are beliefs based on the agent’s percepts (for example, I saw the book on the shelf and I therefore believe it is on the shelf, or I saw the water on the road on a hot summer’s day but I believe this was a mirage as there was no water about); and *generic beliefs* which represent generic rules that the agent believes in (for example, “all men are mortal”, or “there exists birds that do not fly”). However, we do not dwell on the differences between these here, and we simply state that an agent has a set of beliefs: \mathcal{B} .

Desires

Desires are essentially experiences that one desires to take part in. In other words, the percepts composing these experiences are imagined ones, as they have not yet happened or may never happen. For example, one can desire to win the Nobel prize. A desire is then a basic experience \bar{E} such that $\text{content}(\bar{E}) \subset \mathcal{I}$, where \mathcal{I} describes the set of imagined percepts. We note that desires may or may not be feasible. Furthermore, agents may or may not commit to realising their desires; they usually attempt to realise a subset of their desires. Desires that the agent commits to realising through a concrete plan are called intentions, which we introduce next.

Intentions

Intentions are desires with plans: $\text{Intention} = \text{Desire} \times \text{Plan}$, where a plan is defined as a partial order of imagined basic experiences: $\text{Plan} \subseteq \bar{\mathcal{E}} \times \bar{\mathcal{E}}$. We note that basic experiences include basic actions, and an imagined basic action is essentially an action that the agent has committed to realise but has not realised yet. When performing actions, the imagined action will eventually be observed and recorded as an experience with real percepts, as opposed to imagined ones.

Agent Model

An agent is a compilation of its own percepts, experiences, beliefs, desires, and intentions; its capabilities to manifest these formulae; and its capability to perform other actions (such as making commitments, sending messages, and so on). As such, an agent model is then defined as the tuple:

$$\langle \mathcal{P}, \mathcal{E}, \mathcal{B}, \mathcal{D}, \mathcal{I}, \mathcal{C}, \mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{C}, \rangle$$

where \mathcal{P} is the agent's set of all percepts, \mathcal{E} its set of experiences and \mathcal{C} its capability (defined as a function) of manifesting experiences, \mathcal{B} its set of beliefs and \mathcal{B} its capability of manifesting beliefs, \mathcal{D} its set of desires and \mathcal{D} its capability of manifesting desires, \mathcal{I} its set of intentions and \mathcal{I} its capability of manifesting intentions, and finally, \mathcal{C} its capability of performing other actions.

Not all agents will have all these capabilities, and typically they will have a subset of these, and some with different degrees. As such, one may define different categories of agents based on their different capabilities: the degree of capability based (based on \mathcal{C}), of intentionality (based on \mathcal{C} , \mathcal{B} , \mathcal{D} , and \mathcal{I}), and so on. Our research focuses on the intentional agent.

Experience-Based BDI logic

We now define a BDI logic that is grounded on the notion of experience which we will call X-BDI. We first define its syntax, then we give its semantics based on the formal definitions of the previous section, and finally present a sample of the logic's inference rules.

Syntax

- If φ is a propositional well formed formula then $\varphi \in \text{X-BDI}$
- If φ is a propositional well formed formula, $\alpha \in A$ is an agent, and P is a set of plans, then $\text{Action}(\alpha, \varphi) \in P$
- If $\varphi \in \text{X-BDI}$, $\alpha \in A$ is an agent, and $p \in 2^P$ is a plan then $E(\alpha, \varphi), B(\alpha, \varphi), D(\alpha, \varphi), I(\alpha, \varphi, p) \in \text{X-BDI}$
- If $\varphi \in \text{X-BDI}$, $G = \{\alpha, \beta, \dots\} \subseteq A$, and $p \in 2^P$ then $CE(G, \varphi), JE(G, \varphi), SE(G, \varphi, p), CD(G, \varphi), JD(G, \varphi), SD(G, \varphi, p), CI(G, \varphi, p), JI(G, \varphi, p), SI(G, \varphi, p) \in \text{X-BDI}$
- if φ and $\psi \in \text{X-BDI}$ then $\neg\varphi, \varphi \vee \psi \in \text{X-BDI}$

The meaning of the symbols are as follows: B , D , and I are the classical BDI symbols, and E stands for Experience. We use the letter C as a prefix to the E , B , D , and I symbols to mean Common; J to mean Joint; and S to mean Shared. Action has the obvious meaning. Thus, the expression:

$SI(\{\alpha, \beta\}, \text{seeBCNplay}, \{\text{Action}(\alpha, tv), \text{Action}(\beta, stadium)\})$

is a literal in X-BDI expressing that α and β have the shared intention to see Barcelona play a football match by α planning to see the match on tv and β going to the stadium.

Semantics

Formally, we understand the model of the world (which is populated by agents) as a Kripke structure where the state of the world and the mind state of agents evolve due to the actions of agents. For instance, if φ is perceived by α (and it becomes part of α 's experiences), then α may decide to believe φ and we can say that the formula $B(\alpha, \varphi)$ is generated. Similarly, we can define the semantics of formulae like $D(\alpha, \varphi)$. We say agents may have beliefs about other agents, more concretely about the desires, goals, and beliefs of other agents. This means that we will give semantics to

nested expressions. For instance, α may have seen an agent β in the environment watching a football match and jumping with joy when Barcelona scores a goal and thus α processes that experience as β 's desire that Barcelona wins, and that can be represented in the logic as $B(\alpha, D(\beta, \text{winBCN}))$.

Thus, given a set of agents $\{\alpha, \beta, \dots\}$ we define an X-BDI semantic model as a pair $\omega = \langle W, A \rangle$ where W is a classical logic model (i.e. interpretations for propositions) and A is a vector of agent models $A = \langle \alpha, \beta, \dots \rangle$.

Propositions: The interpretation of classical formulae is straightforward. We will use the symbol \models_{PL} to refer to classical satisfaction.

$$\langle W, A \rangle \models \varphi \text{ iff } \varphi \in PL \text{ and } W \models_{PL} \varphi$$

Actions: Differently from other approaches, actions have a *collective* interpretation in our logic, if one agent in the community has a percept of the action being carried on and the world is consistent with the changes in the environment caused by the action then the action is true (recall that \mathcal{AC}_α is α 's set of perceived actions). The satisfaction of a plan consists of the satisfaction of each individual action in the plan.

$$\langle W, A \rangle \models \text{Action}(\alpha, \varphi) \text{ iff } \exists \beta \in A \cdot (\alpha, \varphi) \in \mathcal{AC}_\beta \text{ and } \langle W, A \rangle \models \varphi$$

$$\langle W, A \rangle \models p \text{ iff } p \in 2^P \text{ and } \forall a \in p \cdot \langle W, A \rangle \models a$$

Experiences: A model satisfies that α had an experience φ if it exists in its repository of experiences \mathcal{E}_α . A group of agents had a common experience tagged as λ if all had an experience that semantically entails λ . Similarly, a group of agents had a joint experience tagged as λ if all had an experience that semantically entails λ and they recognise each other as part of the experience. Finally, a group of agents had a shared experience tagged as λ if it was a joint experience and furthermore they all actively participated in it.¹

$$\langle W, A \rangle \models E(\alpha, \varphi) \text{ iff } \varphi \in \mathcal{E}_\alpha$$

$$\langle W, A \rangle \models CE(G, \lambda) \text{ iff}$$

$$\forall \alpha \in G \cdot \exists \varphi \in \text{X-BDI}$$

$$\langle W, A \rangle \models E(\alpha, \varphi) \wedge \varphi \rightarrow \lambda$$

$$\langle W, A \rangle \models JE(G, \lambda) \text{ iff}$$

$$\forall \alpha, \beta \in G \cdot \exists \varphi \in \text{X-BDI} \cdot$$

$$\langle W, A \rangle \models E(\alpha, \varphi) \wedge \varphi \rightarrow \lambda \wedge B(\beta, E(\alpha, \varphi))$$

$$\langle W, A \rangle \models SE(G, \lambda, p) \text{ iff}$$

$$\langle W, A \rangle \models JE(G, \lambda) \text{ and } \langle W, A \rangle \models p \text{ and}$$

$$\forall \alpha \in G \cdot \exists \text{Action}(\alpha, \varphi) \in p \cdot \langle W, A \rangle \models \varphi \rightarrow \lambda$$

Beliefs: The semantics for beliefs is similar to experiences. However, the notion of shared belief is defined to exist when the agents in the group *recognise* one another in a shared experience that led all of them to hold the belief.

$$\langle W, A \rangle \models B(\alpha, \varphi) \text{ iff } \varphi \in \mathcal{B}_\alpha$$

$$\langle W, A \rangle \models CB(G, \varphi) \text{ iff}$$

¹We simplify matters by abusing notation and writing $a \in p$ to mean that action a is 'part' of plan p and $p' \subset p$ to mean that plan p' is 'part' of plan p .

$$\begin{aligned}
& \forall \alpha \in G \cdot \langle W, A \rangle \models B(\alpha, \varphi) \\
\langle W, A \rangle & \models JB(G, \varphi) \text{ iff} \\
& \forall \alpha, \beta \in G \cdot \langle W, A \rangle \models B(\alpha, \varphi) \wedge B(\beta, B(\alpha, \varphi)) \\
\langle W, A \rangle & \models SB(G, \varphi, p) \text{ iff} \\
& \langle W, A \rangle \models JB(G, \varphi) \text{ and} \\
& \exists \lambda \in X\text{-BDI} \cdot \langle W, A \rangle \models SE(G, \lambda, p) \wedge \lambda \rightarrow \varphi
\end{aligned}$$

The same approach used for beliefs can be followed for Desires, which we skip due to lack of space. Next we give semantics for intentions.

Intentions: The intention of an agent to reach a goal by following a plan is true if it is internalised in the intentions of the agents. Common intentions and joint intentions follow the same patterns as in the case of beliefs. However, a shared intention requires that each agent actually intends part of the plan to achieve the goal. That is, no shared intention can happen if any agent is passive in the achievement of the goal.

$$\begin{aligned}
\langle W, A \rangle & \models I(\alpha, \varphi, p) \text{ iff } (\varphi, p) \in \mathcal{I}_\alpha \\
\langle W, A \rangle & \models CI(G, \varphi, p) \text{ iff} \\
& \forall \alpha \in G \cdot \langle W, A \rangle \models I(\alpha, \varphi, p) \\
\langle W, A \rangle & \models JI(G, \varphi, p) \text{ iff} \\
& \forall \alpha, \beta \in G \cdot \langle W, A \rangle \models I(\alpha, \varphi, p) \wedge B(\beta, I(\alpha, \varphi, p)) \\
\langle W, A \rangle & \models SI(G, \varphi, p) \text{ iff} \\
& \forall \alpha, \beta \in G \cdot \exists p_\alpha \subset p \cdot p_\alpha \neq \emptyset \wedge \\
& \langle W, A \rangle \models I(\alpha, \varphi, p_\alpha) \wedge B(\beta, I(\alpha, \varphi, p_\alpha))
\end{aligned}$$

Inference

We assume classical propositional deduction for propositional fragments of X-BDI and modus ponens on the X-BDI formulae. X-BDI axioms are those of propositional logic plus specific axioms for X-BDI expressions. Due to space limitations, here we include axioms for Desires only:

$$\begin{aligned}
& \vdash D(\alpha, \varphi) \wedge D(\beta, \varphi) \leftrightarrow CD(\{\alpha, \beta\}, \varphi) & (1) \\
& \vdash CD(G, \varphi) \wedge D(\beta, \varphi) \leftrightarrow CD(G \cup \{\beta\}, \varphi) & (2) \\
& \vdash JD(G, \varphi) \rightarrow CD(G, \varphi) & (3) \\
& \vdash CD(\{\alpha, \beta\}, \varphi) \wedge B(\alpha, D(\beta, \varphi)) \wedge \\
& \quad B(\beta, D(\alpha, \varphi)) \leftrightarrow JD(\{\alpha, \beta\}, \varphi) & (4) \\
& \vdash JD(G \cup G', \varphi) \rightarrow JD(G, \varphi) \wedge JD(G', \varphi) & (5) \\
& \vdash SD(G, \varphi, p) \rightarrow JD(G, \varphi, p) & (6) \\
& \vdash JD(G, \varphi, p) \wedge SE(G, \lambda, p) \wedge (\lambda \rightarrow \varphi) \rightarrow \\
& \quad SD(G, \varphi, p) & (7)
\end{aligned}$$

Example

With the basic definitions we can now specify the exact steps needed for intentional agents to achieve shared intentions (and hence, shared experiences).

Investigating & Realising a joint desire. Agents may have different motives for engaging in shared experiences, based on their own personal beliefs and desires. However, when one agent realises that it would like to fulfil its desire

of having a shared experience with other agents, then it will contact those agents in the hope that it can convince them to share its desire, and eventually fulfil that desire. In multi-agent systems, this could be achieved through argumentation. This step describes the agent's investigation of whether a joint desire for achieving the shared experience may be realised. If agents agree that they are all interested in realising a shared experience, then a joint desire to achieve the shared experience is born (i.e. it becomes true).

Investigating & Realising a shared intention. After a joint desire is born, agents are then committed to finding the plan of action for fulfilling the desired shared experience. Argumentation is usually used here to make sure that all parties agree to who does what and under what conditions. If the agents succeed in agreeing on a plan of action, then the shared intention for executing the plan for achieving the desired shared experience is now born (i.e. it becomes true).

Realising a shared experience. Agents are now committed to carrying out the actions of the plan that they are responsible for. When the plan has been executed, the shared experience is realised.

In what follows, we take the scenario of two agents that are interested in the shared experience of buying a gift together. Let us say the agents adopt the names of their human owners, c for Carla and b Bill, and they are interested in buying a gift for Mary together (the desire to buy a gift for Mary is referred to as g). In what follows, we describe what actions could the agents perform based on their X-BDI, and the resulting change in their X-BDI for each action of the environment.

0. **Motivating the interest in a shared experience.** Say Carla is interested in buying Mary a gift, but because she cannot afford a gift on her own, she desires to buy the gift with someone else (referred to as *anyone*). As such, her belief base includes:

$$D(c, SE(\{c, anyone\}, g, inanyway)) \quad (i)$$

Note that *inanyway* refers to some plan of action that has not been decided upon yet.

Bill, on the other hand, desires to share any experience with Carla (referred to as *anything*), either because he is dependent on her, or possibly secretly in love with her. As such, Bill's belief base contains the following:

$$D(b, SE(\{c, b\}, anything, inanyway)) \quad (ii)$$

1. **Investigating & Realising a joint desire.** Carla's desire to find a partner for buying Mary a gift with (literal (i) in Carla's belief base) drives her to contact Mary's friends hoping to find a match, asking each "Would you like to buy a gift for Mary with me?" Moreover, Bill's desire to share an experience with Carla (literal (ii) in Bill's belief base) drives him to reply with a "Yes". This exchange of information leads to Carla modifying her belief base to contain the following:

$$B(c, D(b, SE(\{c, b\}, g, inanyway))) \quad (iii)$$

$$B(c, B(b, D(c, SE(\{c, b\}, g, inanyway))))$$

And deducing that:

$$B(c, JD(\{c, b\}, SE(\{c, b\}, g, inanyway)))$$

Similarly, Bill modifies his belief base by adding:

$$B(b, D(c, SE(\{c, b\}, g, \text{in anyway}))) \quad (iv)$$

$$B(b, B(c, D(b, SE(\{c, b\}, g, \text{in anyway}))))$$

And deducing that:

$$B(b, JD(\{c, b\}, SE(\{c, b\}, g, \text{in anyway})))$$

As such, the following now holds:

$$JD(\{c, b\}, SE(\{c, b\}, g, \text{in anyway})) \quad (v)$$

Note that the joint desire for a shared experience may be the result of various individual desires. For instance, Carla just needs someone to share her expenses, whereas Bill just want to do anything with Carla.

Also note that (v) is deduced from axiom (4), the belief literals (iii) and (iv), and formula $CD(\{c, b\}, SE(\{c, b\}, g, \text{in anyway}))$, which in turn is deduced from axiom (1) and belief literals (i) and (ii).

All the deduced formulae that we mention in this section are deduced in a similar straightforward manner. However, due to space limitations, and since we do not provide the exhaustive set of inference rules, we have only provided (above) the proof for formula (v).

2. **Investigating & Realising a shared intention.** After both agents agree that they share the desire of buying the gift together, they go on to argue on the details of the plan they will follow. For example, where will they buy the gift from? What should they buy Mary? And so on. In this paper, we do not dwell on the details of argumentation. However, assuming both agents agree on a plan p , and they both communicate their agreement to each other, then Carla's belief base is updated to contain the following:

$$I(c, SE(\{c, b\}, g, p), p_c)$$

$$B(c, I(b, SE(\{c, b\}, g, p), p_b))$$

$$B(c, p_c \subset p \wedge p_c \neq \emptyset \wedge p_b \subset p \wedge p_b \neq \emptyset)$$

And the following is then deduced:

$$B(c, SI(\{c, b\}, SE(\{c, b\}, g, p), p))$$

Note that p_c is Carla's part of the plan and p_b is Bill's part of the plan.

Similarly, Bill's belief base is updated by adding:

$$I(b, SE(\{c, b\}, g, p), p_b)$$

$$B(b, I(c, SE(\{c, b\}, g, p), p_c))$$

$$B(b, p_c \subset p \wedge p_c \neq \emptyset \wedge p_b \subset p \wedge p_b \neq \emptyset)$$

And the following is deduced:

$$B(b, SI(\{c, b\}, SE(\{c, b\}, g, p), p))$$

As such, the following now holds:

$$SI(\{c, b\}, SE(\{c, b\}, g, p), p)$$

Note that each agent commits to performing its own part of the plan (in this case, p_c and p_b , respectively) to make sure the plan is realised, and both agents believe that this plan will realise their shared experience.

3. **Realising a shared experience.** After executing their actions and the plan is fulfilled, if the agents can perceive that the plan has been fulfilled, then Carla's belief base is updated to contain the following:

$$B(c, SE(\{c, b\}, g, p))$$

Similarly, Bill's belief base is updated by adding:

$$B(b, SE(\{c, b\}, g, p))$$

And the shared experience is said to have been realised:

$$SE(\{c, b\}, g, p)$$

Background

In our literature review, we do not focus on BDI approaches in multiagent system (such as the work of Bordini, Hübner, and Wooldridge (2007); Winikoff (2005); Rao and Georgeff (1991); Casali, Godo, and Sierra (2011); Rao (1996); Dastani (2008); Parsosn and Giorgini (1999)) because to our knowledge, there is no mention of the notion of experience and shared intentionality in current BDI approaches, which is the main contribution of our work. Individual actions are dictated by individual intentions, which are the result of individual beliefs, and desires. But how do collective actions come about? In this section, we relate our work to existing philosophical views that have discussed these issues (as summarised by Tollefsen (2004)). Philosophers, especially those interested in action theory, have been more and more interested in the notion of collective actions, intentionality, and belief.

Some argue that the collective attitude simply refers to the fact that the majority of the group's members share that attitude (Quinton, 1975). These accounts were labelled as *summative* accounts by Gilbert (1987, 1989). For instance, simple summative accounts (SSA) state that: *Group G intends p if and only if all or most of the members intend p*. However, it has been argued that the SSA is not sufficient (e.g. Gilbert (1987, 1989)). The complex summative accounts (CSA) try to address this problem by introducing the notion of common knowledge. The CSA states that: *Group G intends p if and only if: (1) most of the members of G intend p, and (2) it is common knowledge in G that (1)*.

In our model, simple summative accounts are described by the *common* modalities. For example, a common experience is nothing more than an experience that has been experienced by the individuals of the group. Similarly, a common belief is a belief that happens to be shared by the group's members. Common desires and intentions are defined similarly. Complex summative accounts are described by the *joint* modalities. For example, a joint experience is an experience that is common to a group, and the members of the group are aware that they have all experienced the same experience. Similarly, a joint belief is a belief that is common to the members of the group, and the members are aware that they share the same belief. Again, joint desires and intentions are defined similarly.

Gilbert argues that both the SSA and CSA accounts are not the right approaches since a group's attitude cannot always be described in terms of the individual ones, even if it was enhanced with common knowledge (Gilbert, 1989, 1994). For instance, a group of people who are jogging in the morning in the park are aware of each other's intentions to jog; yet, there is something different between this group of people and a group of friends who decide to jog together.

While our model does define common and joint modalities, we note that these collective modalities are simply used to describe the collection of individual modalities. They are not used to describe the groups' intentions as a whole. For that, the *shared* modality is introduced. A shared experience is a joint experience in which the agents are actively involved. A shared intention is an intention shared by the group members to carry out a predefined plan for fulfilling a

joint desire. However, before we dwell on the shared modalities, we first introduce the philosophers' different views, and then compare them to our proposed model.

Searle (1990, 1995) states that the we-intend cannot be reduced to a set of I-intend, even if it was supplemented with mutual beliefs. For example, having two strangers walking down the street that happen to walk next to each other is different than having a couple walking *together* down the street. As such, Searle argues that collective intentions should combine the sense of acting with the sense of willing something together. In our model, the notion of shared experience does not necessarily imply that the different agents sharing an experience were willing to perform the actions they did together. For example, a couple of strangers who intend to jog in the morning may share this experience if they happen to be at the same place at the same time. However, the shared intention is defined as having a shared plan, where each agent intends to fulfil its part of the plan. The very definition of an agent's intention is its willingness and commitment to execute the corresponding plan. As such, a shared intention is realised only if the agents are willing and committing to execute their actions as detailed by the plan agreed upon. An example of this would be a couple who agree to go jogging together in the morning.

Searle has been strongly criticised for his "brain in a vat condition", which states that intentionality could be held by a brain in a vat. We do not study how beliefs are formed and whether an agent can form a belief about something it has never encountered, or if two agents can share a belief if there is no alignment between their ontologies. This is outside the scope of our work. However, we do note that although a joint intention can either be true or false, an agent may still hold false beliefs about a joint intention. This happens if it holds false beliefs about other agents' beliefs and intentions.

Bratman (1992, 1999), like Searle, does not believe in a plural agent that could hold shared intentions, nor does he believe that shared intentions can be reduced to individual intentional states. In his discussion, Bratman uses the word "shared intention" as opposed to "collective intention" to refer to interrelated individual intentional states. In his definition, an intention is shared if and only if our intentional actions are coordinated by making sure our personal plans of action meld together. As such, shared intentions give rise to argumentation and negotiation for agreeing on the coordinated plans of actions. Bratman further argues, like Searle, that a single agent can have a we-intention (Bratman, 1997, 1999), which has attracted a great deal of criticism (Stoutland, 1997; Velleman, 1997), since one cannot intend what he/she cannot fulfil (in other words, one cannot intend actions that may only be executed by others).

Similar to Bratman, we say an intention is shared only if there is a shared plan of action that meshes and coordinates individual actions. The level of coordination, however, remains loose. For example, the plan may simply state who does what, without going into the order or the pre and post conditions of actions. However, unlike Bratman and Searle, we say a shared intention cannot be held by a single agent (although single agents may form beliefs about the truth of shared intentions), but by the group of agents who the plan

of action cannot be carried out without. Agents can only intend to perform their part of a shared plan.

Finally, we note that Bratman refused to accept the notion of obligations and promises in shared intentions, arguing that promises are neither sufficient nor necessary for shared intentions. Gilbert (1989, 1996) argued otherwise, and as such, believed that both Searle and Bratman have failed to consider the normative element of shared intentions, or the resulting commitments of a shared plan. For instance, if one of the couple who decided to jog together in the morning decides to leave in the middle of the walk, this will be considered as a violation of a commitment, and the other has the right to take offense. In our model, individual intentions are defined as the agent's commitment to execute the predefined plan for achieving a given goal. As such, a shared intention becomes the commitment of each member to the shared plan for achieving a given goal. However, there are no social norms attached to shared intentions that specify how breaching a certain commitment is dealt with, or how agents lying about their intentions are punished. This does not contradict the proposed model, but lies outside its scope.

Conclusion

In this paper, we have provided a novel and formal agent model by introducing the notion of experience, which situates agents in their environment, providing a clear and tangible link on how the environment (and its changes) impact the evolution of an agent's BDI. Experiences help complete the BDI cycle: experiences shape beliefs (usually resulting in belief updates), beliefs influence desires, and desires drive intentions. Intentions in turn lead the agent to act in its environment (where the action could be as simple as to observe), which results in the agent having new experiences.

We have also built a logical formalism, the X-BDI, to describe agent reasoning whose semantics are based on our proposed agent model. Another novel contribution of this work is defining, through the X-BDI logic, the notion of shared experience and using it as a fundamental construct to underpin collective intentionality. Our work essentially motivates shared experiences and intentionality: the basis of social behaviour and the foundation of multiagent systems in general.

Our ongoing work provides a blueprint for an experience-based BDI agent architecture, specified in the Z notation (Spivey, 1989). This means that the semantics of the logic can much more readily be used for the implementation of agent systems and so we address arguably the most significant issue with new logics of agency that there is no clear link between the theory they describe and the practice of building real systems.

Our future plans are to formally prove the soundness and completeness of the inference system and to develop a P2P agent system developed from the Z specification of our agent architecture along with the X-BDI reasoning system. We are using the theory outlined here to build social browsers that enable human users, supported by software agents, to share online cultural experiences such as jointly visiting the website of a museum from their homes.

References

- Bordini, R. H.; Hübner, J. F.; and Wooldridge, M. 2007. *Index*. John Wiley & Sons, Ltd. 269–273.
- Bratman, M. E. 1992. Shared cooperative activity. *Philosophical Review* 101(2):327–341.
- Bratman, M. E. 1997. I intend that we J. In Holmström-Hintikka, G., and Tuomela, R., eds., *Contemporary Action Theory Volume 2: Social Action*, volume 267 of *Synthese Library*. Dordrecht, Netherlands: Kluwer Academic Publishers Group. 49–64.
- Bratman, M. E. 1999. *Faces of Intention*. Cambridge Studies in Philosophy. Cambridge University Press.
- Casali, A.; Godo, L.; and Sierra, C. 2011. A graded bdi agent model to represent and reason about preferences. *Artif. Intell.* 175(7-8):1468–1478.
- Dastani, M. 2008. 2apl: a practical agent programming language. *Autonomous Agents and Multi-Agent Systems* 16(3):214–248.
- d’Inverno, M.; Luck, M.; Georgeff, M. P.; Kinny, D.; and Wooldridge, M. 2004. The dMARS architecture: A specification of the distributed multi-agent reasoning system. *Autonomous Agents and Multi-Agent Systems* 9(1–2):5–53.
- Ekman, P. 1972. Universals and cultural differences in facial expression of emotion. In Cole, J., ed., *Nebraska Symposium on Motivation*, volume 19, 207–283. Lincoln: University of Nebraska Press.
- Gilbert, M. P. 1987. Modelling collective belief. *Synthese* 73:185–204. 10.1007/BF00485446.
- Gilbert, M. P. 1989. *On social facts*. International library of philosophy. New York, NY: Routledge.
- Gilbert, M. P. 1994. Remarks on collective belief. In Schmitt, F. F., ed., *Socializing Epistemology: The Social Dimensions of Knowledge*, Studies in epistemology and cognitive theory. Lanham, Maryland: Rowman and Littlefield Publishers. 235–56.
- Gilbert, M. P. 1996. Concerning sociality: The plural subject as paradigm. In Greenwood, J. D., ed., *The Mark of the Social: Discovery or Invention?* Lanham, Maryland: Rowman and Littlefield Publishers. 17–36.
- Kant, I. 1838. *Critick of Pure Reason*. W. Pickering: London.
- Kitcher, P. 1990. *Kant’s Transcendental Psychology*. Oxford University Press.
- Parsosn, S., and Giorgini, P. 1999. *An approach to using degrees of belief in BDI agents*. Kluwer. 81–92.
- Perrett, D. I.; Xiao, D.; Barraclough, N. E.; Keysers, C.; and Oram, M. W. 2009. Seeing the future: Natural image sequences produce “anticipatory” neuronal activity and bias perceptual report. *The Quarterly Journal of Experimental Psychology* 62(11):2081–2104.
- Quinton, A. 1975. The presidential address: Social objects. *Proceedings of the Aristotelian Society* 76:1–27+viii.
- Rao, A. S., and Georgeff, M. P. 1991. Modeling rational agents within a bdi-architecture. In *KR*, 473–484.
- Rao, A. S. 1996. Agentspeak(1): Bdi agents speak out in a logical computable language. In *MAAMAW*, 42–55.
- Schütz-Bosbach, S., and Prinz, W. 2007. Prospective coding in event representation. *Cognitive Processing* 8(2):93–102.
- Searle, J. R. 1990. Collective intentions and actions. In Cohen, P. R.; Morgan, J.; and Pollack, M. E., eds., *Intentions in Communication*, Bradford Books. Cambridge, MA: MIT Press. 401–415.
- Searle, J. R. 1995. *The construction of social reality*. Free Press.
- Spivey, J. M. 1989. *The Z notation: a reference manual*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
- Stoutland, F. 1997. Why are philosophers of action so anti-social? In Alanen, L.; Heinämaa, S.; and Wallgren, T., eds., *Commonality and particularity in ethics*, Swansea studies in philosophy. New York, NY: St. Martin’s Press.
- Szpunar, K.; Watson, J.; and McDermott, K. 2007. Neural substrates of envisioning the future. *Proc Natl Acad Sci U S A*.
- Tollefsen, D. 2004. Collective intentionality. In *The Internet Encyclopedia of Philosophy*.
- Velleman, J. D. 1997. How to share an intention. *Philosophy and Phenomenological Research* 57(1):29–50.
- Weiler, J. A.; Suchan, B.; and Daum, I. 2010. When the future becomes the past: Differences in brain activation patterns for episodic memory and episodic future thinking. *Behav Brain Res* 212(2):196–203.
- Winikoff, M. 2005. Jack intelligent agents: An industrial strength platform. In Bordini, R. H.; Dastani, M.; Dix, J.; and Fallah-Seghrouchni, A. E., eds., *Multi-Agent Programming*, volume 15 of *Multiagent Systems, Artificial Societies, and Simulated Organizations*. Springer. 175–193.