Shared Experiences and Intentionality

Nardine Osman
Mark d’Inverno
Carles Sierra

December, 2011

Technical Report IIIA 2011-07

Campus de la Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona
Shared Experiences and Intentionality


Nardine Osman* and Mark d’Inverno* and Carles Sierra
Artificial Intelligence Research Institute (IIIA-CSIC), Barcelona, Spain
Computing Department, Goldsmiths, University of London

December 20, 2011

Abstract
Introducing the notion of experience as a primitive construct, we illustrate how the BDI cycle is completed: experiences generate beliefs, which in turn generate desires, which in turn generate new experiences. Furthermore, the notion of experience is used to ground social BDI agents in the physical world. We do not assume an agent’s BDI model already exists, but we illustrate how the agent’s mind, along with its environment (or its perceptions of the environment, to be more precise), helps shape and evolve its BDI belief base. The notions of shared experience and shared intentionality then naturally arise, helping us better understand and motivate social behaviour: the foundation of multiagent systems.

Introduction
In this work, we do not assume an agent’s BDI model already exists, but we try to explain how it is built through the agents’ capabilities of perceiving the environment, imagining and desiring, manifesting beliefs based on percepts (and possibly desires), and so on. We basically provide a trivial definition of what an environment is, followed by a set of definitions of the required agent’s capabilities for manifesting perceptions, experiences, beliefs, desires, and intentions. In summary, this paper essentially provides a grounding framework for social BDI agents in real environments where an agent’s perception of its environment is formally modelled, resulting in the novel notion of ‘experience’. We then illustrate how this notion of experience helps complete the BDI cycle resulting in a dynamically evolving BDI.

As a consequence, we then build the notion of shared experience on top of the notion of single individual experiences. This paper’s proposed model helps provide a formal specification, with a computational semantics, of an abstract agent architecture for social BDI agents. We argue that such a model motivates social behaviour. For instance, it helps explain why would one agent approach another agent; why do agents speak in the first place. Furthermore, shared experiences may be used as the bases for shared intentionality, which drives group behaviour, as opposed to individual intentionality, which drives individual behaviour.

Last, but not least, the framework can then be used to design agent-based technologies for enabling shared social experiences in human societies. Naturally, social BDI agents can then be designed and tested to evaluate the claim that social agents produce a more flexible and robust society. With such a framework, one can also evaluate the impact of technology-mediated shared experiences, which we currently leave for future work. This can lead to increased refinements of our model but also provide an evaluation of its ability to model social human behaviour.

The remainder of this paper is divided as follows: it first starts with a trivial definition of the physical world, followed by a more comprehensive definition of the agent’s mental world (which provides the grounding for the logic’s computational semantics), then the logic for the experience based BDI model (X-BDI) is introduced, followed by a motivating example that illustrates the impact of this work, some literature review of related work, and finally, a concluding statement.

The Physical World
We say the physical world is composed of objects that populate the environment and events that occur in the environment, usually resulting in changing the environment. Naturally, some of these objects may be viewed as agents, but this depends on the perspective of the observer. For instance, for some, an agent is an object with a goal. For example, a cup may be temporarily labelled as an agent as long as it holds the goal of holding the water. For others, an agent is an intentional agent that acts based on intentions. However, we leave this categorisation outside the definition of the physical world since we assume this categorisation is always held by an observer and is not encoded in the physical world itself. Similarly, some events may be thought of as actions performed by an actor, or an agent. Again, this categorisation is held by the observer and not encoded in the physical world. As such, we say the physical world is defined by a set of objects and events. Some of these objects may be labelled as intentional agents. These are the agents that the rest of this paper focuses on. Furthermore, these agents are capable of labelling

Last, but not least, both objects and events have properties. For instance, the temperature, the color, the mass are
all properties of objects. Spacial and temporal information may be properties of both objects and events. In this paper we do not go into the details of these properties and their categorisation in the physical world, as this is a very philosophical issue. However, we do define the agents cognitive skills in perceiving and representing these properties.

As such, the physical world in our model is defined in the following straightforward manner, which simply requires a set of objects and a set of agents, both of which may hold properties.

\[ \text{Object}, \text{Event} \]
\[ \text{Property} \]

\[ \text{ObjectEntity} \equiv \text{Object} \times \mathbb{P} \text{Property} \]
\[ \text{EventEntity} \equiv \text{Event} \times \mathbb{P} \text{Property} \]

\[ \text{Entity} \equiv \text{object} \langle \text{ObjectEntity} \rangle \mid \text{event} \langle \text{EventEntity} \rangle \]

\[ \text{Environment} \equiv \mathbb{P} \text{Entity} \]

**The Mental World**

**Perception**

We say an agent is capable of perceiving its environment, and manifesting a set of percepts. Observing the immediate physical world is a main source of one’s percepts. However, we also argue that one may imagine a percept, where the imagination is usually based on other observed percepts. For example, imagining one can fly is not necessarily based on having perceived one can fly in the past, but on perceiving flight in general and imagining the application of flight to oneself. This follows the cognitive neuroscience view that considers imagination as a second order perception. (?) We call these the observed and imagined percepts, and we define the set of these percepts accordingly:

\[ \text{[ObservedPercept}, \text{ImaginedPercept}] \]

A percept in general is then defined accordingly:

\[ \text{Percept} \equiv \text{obs} \langle \text{ObservedPercept} \rangle \mid \text{imagine} \langle \text{ImaginedPercept} \rangle \]

**Manifesting Perceptions**

We assume agents have distinct capabilities for manifesting percepts of the above two different types. Specifically, two different functions, perceive and imagine, describing the agent’s capability to observe its environment and imagine non-existing ones. However, we note that when observing the environment, the resulting perception is shaped by both the observed environment and the observer. For instance, the color of a wall (which exists in the environment) might be perceived differently by different observers. This could either be due to physical constraints (such as the luminosity), due to the difference in the sensory receptors of the observers (consider, for instance, colour blind people), or even due to the different mindset of the observer, amongst other things. Neuroscientists argue that there are no “absolute truth” perceived by observers. Each perception is the result of one’s sensory receptors plus the interpretation of the observer (where the interpretation is defined through the observer’s current *mind set*). (?)

Before we define this “interpretation” or “mindset”, we follow the view (?) that says that an agent’s memory is divided into three main parts: (1) sensory memory, which is short term and simply holds the percepts that are currently being perceived by the agent’s sensory receptors; (2) working memory, which describes what the agent is currently conscious about (such as the currently active percepts, experiences, beliefs, desires, etc.); and (3) long term memory, which holds the agent’s entire history of percepts, experiences, beliefs, etc. The sets of sensory, working, and long term memories are defined accordingly.

\[ \text{[SensoryMemory}, \text{WorkingMemory}, \text{LongTermMemory}] \]

An agent’s memory then is a compilation of all three.

\[ Memory \equiv \text{sensory} \langle \text{SensoryMemory} \rangle \mid \text{working} \langle \text{WorkingMemory} \rangle \mid \text{longterm} \langle \text{LongTermMemory} \rangle \]

We argue that sensory memory usually does not map to the “true” state of the environment, as the degree of what may be perceived may change from one receptor to another, while others receptors may simply be defective. We state that sensory memory is automatically registered in the working memory, although the working memory (which defines the agent’s “interpretation”, or “mindset”) will impact how sensory memory is being registered there. We also argue that everything that is added to the working memory is automatically registered to the long term memory. In other words, the working memory is always a subset of the long term memory. However, in the remainder of this paper, we simplify things by simply referring to the agent’s entire memory, instead of going into detail of which part of this memory will impact the agent’s various capabilities (such as manifesting perceptions).

As such, the functions perceive and imagine will require the presence of a memory, as well as the environment in which the agent is situated.

| environment : Environment |
An agent’s capability to manifest percepts is then defined accordingly.

\[ \text{ManifestingPerceptions} \]
\[ perceive : (\text{Environment} \times \text{Memory}) \to (\mathbb{P} \text{ObservedPercept} \times \text{Memory}) \]
\[ imagine : \text{Memory} \to (\mathbb{P} \text{ImaginedPercept} \times \text{Memory}) \]

\[ \forall o : \mathbb{P} \text{ObservedPercept}; m_1, m_2 : \text{Memory}; e : \text{Environment} \cdot \]
\[ \text{perceive}(e, m_1) = (o \times m_2) \Rightarrow (o \not\subseteq m_1 \land o \subseteq m_2) \]

\[ \forall i : \mathbb{P} \text{ImaginedPercept}; m_1, m_2 : \text{Memory} \cdot \]
\[ \text{imagine}(m_1) = (i \times m_2) \Rightarrow (i \not\subseteq m_1 \land i \subseteq m_2) \]

where the conditions state that the resulting percepts did not exist in the old memory and that they are added to the new memory.
Recalling Percepts, amongst other things  We have noted earlier that sensory memory only contains percepts, and we have discussed how these percepts may be introduced to the sensory memory and their registration in the working memory and the long term memory. In summary, we have discussed the flow of percepts “downwards”. But how do percepts (or other elements of the long term memory, such as experiences, beliefs, desires, etc.) flow “upwards”? This is crucial for specifying what the agent is currently conscious about. For this, we define a recall function that helps the agent recall elements of its long term memory and registering them in the working memory. However, when registering something in the working memory, whether the source was either the sensory memory or the long term memory, this registration is influenced by the working memory itself (also referred to as the agent’s current interpretation, or mindset). That is because the agent recalling a percept has a different mindset when recalling it from when perceiving or imagining it. This new mindset results in a new interpretation, and hence, a new perception. For example, a percept (or an experience) may change in exaggeration or to suit ones view of oneself. Discussing past experiences will usually modify those experiences, where the modification may be represented through the recall function. Again, this view is in accordance with that of the neuroscientists, who assert that recalling a perception always results in overwriting the original one based on the recalling agent’s current interpretation. (1) In other words, when a percept (or another element of the long term memory) is recalled, it is usually modified when it is registered in the working memory, which results in overwriting it in the long term memory (remember that the working memory is always a subset of the long term one). We do acknowledge that in some cases of artificial agents, an agent’s recall function could be impartial to the agent’s mindset. In such cases, a simplification may be made, and recalling percepts will not result in their modification: recalling would simply be accessing existing percepts. We note that the recall function may also represent the forgetfulness (or memory loss) of agents when it fails to recall certain elements of the long term memory. An agent’s capability to recall elements of its long term memory is then defined accordingly.

\[
\forall m_1, m_2 : \text{Memory} \bullet \text{recall}(m_1) = m_2 \Rightarrow \\
\exists m'_1 \subseteq m_1 \land m'_2 \subseteq m_2 \bullet m'_1 \neq \{\} \land m'_2 \neq \{\} \land \\
(m'_1 \cup (m_1 \setminus m'_2)) \oplus (m'_2 \cup (m_2 \setminus m'_1))
\]

where the condition above states that the old memory is modified by overwriting some of its elements.

We remind the reader that although we present the recall function under the section that discusses percepts, this function may be applied to recall any element of the agent’s long term memory, such as experiences, beliefs, desires, and intentions, which we present shortly.

Types of Percepts  The environment is composed of objects with properties, some of which may be labelled as agents. An example of these entities are “my tall neighbour”, “the green shirt”, and so on. Agents should be capable of perceiving (some of) these entities and their properties, based on their sensory receptors’ capabilities. As such, “my tall neighbour” or “the green shirt” would be represented as percepts manifested by the agent.

\[
\text{objectPercepts} : \mathbb{P} \text{Percept} \\
\text{agentPercepts} : \mathbb{P} \text{Percept}
\]

We note that of particular interest to artificial intelligence are agents and their actions in the environment, since these are crucial in reasoning and planning. As such, distinguishing between objects and agents is an important capability of intentional agents. We also note that a percept is not identical to the object in the real world. The percept represents the agent’s perception of something. For example, “my neighbour” is an agent percept that corresponds to the object “Claire” in the environment.

In addition, some agents are also capable of perceiving some events in their environment, such as “rainfall”, “sunset”, and so on. Furthermore, some agents are also capable of distinguishing which events are performed by other agents, such as “Mark is singing”, “that girl is dancing”, “I am getting angry”, and so on. We refer to these as the event and action percepts, respectively.

\[
\text{eventPercepts} : \mathbb{P} \text{Percept} \\
\text{actionPercepts} : \mathbb{P} \text{Percept}
\]

We note that event and action percepts usually map to events and actions that take place in the environment.

Naturally, these different types of percepts are mutually disjoint.

\[
\text{disjoint}\{\text{objectPercepts}, \text{agentPercepts}, \\
\text{eventPercepts}, \text{actionPercepts}\}
\]

Finally, we remind the reader that object, agent, event, and action percepts can be either observed or imagined.

Symbols  But how can one analyse percepts and communicate them? For instance, given the percept “the blue shirt”, how can one answer questions such as “What is the color of the shirt?” And how can one communicate their percepts to others. These are essentially linguistic issues, and the Ogden/Richards triangle (2), also known as the triangle of reference or the semiotic triangle, is one common way of addressing these issues. Figure 1 presents the semiotic triangle which describes three basic entities: the referent, the thought (or reference), and the symbol. The basic idea is that the referent represents something that exists in the real world (or the environment), the thought refers to the abstract concept in the mental mind which refers to the referent, and the symbol is used to symbolise the thought and present communication means that would help discuss a thought or a referent.
Note that the link between a symbol and a referent is a deduced one: a symbol cannot directly refer to a referent, but indirectly refer to it through the corresponding abstract concept.

Figure 1: The Ogden/Richards triangle

This view maps neatly with our proposed model, where we say: the referent is an object in the real world, an agent, an event, or an action; the thought is the percept; and the symbols compose the agent’s language to help it describe and talk about its percepts. As such, the agent will require a set of symbols.

We note, however, that percepts may either be observed or imagined. In the former case, the percept will be linked to an object in the real world, which is not always true in the latter case. As such, we note that not all thoughts (percepts) are linked to the real world. Imagined percepts, for instance, may help one think of abstract objects (such as the concept of a chair), while observed percepts would represent objects in real life (the chair one is currently sitting on).

Manifesting Symbols & Mapping them to Percepts The agent requires a mechanism to generate symbols that would describe its percepts. This is crucial for communication. Similarly, it also needs to relate symbols to their corresponding percepts. We note that the latter is not necessarily the inverse of the former. Furthermore, we note that both functions are not injective as we are assuming two symbols can relate to the same percept, and two percepts can also map to the same symbols. For instance, “my neighbour” and “Claire” can refer to the same percept. Similarly, two different perceptions of cars can both map to the symbol “car”.

The agent’s capabilities to perceive the environment and represent percepts / interpret symbols compose the two basic legs of the semiotic triangle (see Figure 1), where perceive maps to the relation between a referent and a thought (or a percept), and the represent/interpret functions map to the bidirectional relation between a thought and a symbol.

Symbols for Reasoning and Communication We can define a query that allows one to interrogate its own percepts, i.e. reason over its percepts.

Note that the first symbol represents a question over a given percept, and the second represents the answer to the question. For instance, consider the action percept “Claire played the electric piano”. Queries may be used to deduce which agent performed the action by asking “Who played the electric piano?”. Queries may also help answer questions about which event did the agent execute by asking “What did Claire do?”. These types of questions help with reasoning over agents’ actions. For this reason, we distinguish two special types of queries that when applied to action percepts would return the agent percept describing which agent carried out the action or the event percept describing what is event percept that was performed.

<table>
<thead>
<tr>
<th>agent : (Symbol × Percept) → Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>∀ s₁, s₂ : Symbol, p : Percept • agent(s₁, p) = s₂ ⇒ (p ∈ actionPercepts) ⇒ (interpret(s₂) ⊆ agentPercepts)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>event : (Symbol × Percept) → Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>∀ s₁, s₂ : Symbol, p : Percept • event(s₁, p) = s₂ ⇒ (p ∈ actionPercepts) ⇒ (interpret(s₂) ⊆ eventPercepts)</td>
</tr>
</tbody>
</table>

However, although we do distinguish the above two special types of queries, we note that queries could be very general. They may be about the properties of things, such as asking “What was the type of the piano played by Claire?”. Queries may also have binary answers: “yes/no”. For instance, the answer to the query “Did Claire play the piano?” is “Yes”.

As mentioned earlier, symbols are usually related to the percepts. For instance, when answering “Who played the electric piano?”, the answer symbol “Claire” should map to the agent percept that refers to Claire. However, we do note that not all symbols are directly related to percepts. For instance, “yes/no” symbols do not map to specific percepts.

Knowledge Representation & Reasoning The agent’s capability to represent and reason over its own knowledge is then a compilation of the above functions.

<table>
<thead>
<tr>
<th>KnowledgeRepresentationReasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>represent : Percept → Symbol</td>
</tr>
<tr>
<td>interpret : Symbol → Percept</td>
</tr>
<tr>
<td>query : (Symbol × Percept) → Symbol</td>
</tr>
</tbody>
</table>

Experiences In its most basic form, an experience is defined as a number of manifested 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 that the basic cognitive hard wired relationships are time ordering, causality, and spatial
relationships. As such, we define the functions $\prec$, $\sim$, and $\triangleright$ to define those three relations, accordingly. We also adopt Kant’s view that human beings have moral imperative that makes them classify things as good or bad. As such, we introduce the evaluation function $\triangledown$ 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” or “losing made me angry”.

All the above functions are discussed in more detail shortly. However, we note that the spatial relation requires the agent to have a predefined set of spatial relationships that would describe the spatial order between the elements of an experience, which we define accordingly:

$$[\text{SpatialRelationships}]$$

Examples might include above, below, left, right, behind, in front, and so on.

Furthermore, the evaluation function also requires the agent to have a predefined evaluation space that describes the different degrees to which an agent may like or hate an experience, which we define accordingly:

$$[\text{EvaluationSpace}]$$

Examples may include great, good, average, bad, horrible, and so on. Alternatively, the evaluation space may simply be the set of natural numbers.

An experience is then defined as follows:

$$\text{CompoundExperience} \equiv \begin{cases} \text{Symbol} \\ \mathcal{P} \text{Experience} \\ \mathcal{P}(\text{Experience} \times \text{Experience}) \times \mathcal{P}(\text{Experience} \times \text{Experience}) \times \mathcal{P}(\text{Experience} \times \text{Experience} \times \text{SpatialRelationships}) \times \text{optional}(\text{Symbol} \rightarrow \text{EvaluationSpace}) \end{cases}$$

$$\& \text{Experience} \equiv \begin{cases} \text{basic}(\langle \text{Percept} \rangle) | \text{compound}(\langle \text{CompoundExperience} \rangle) \end{cases}$$

The following is used to label the different elements of a compound experience:

$$\begin{array}{|l|} \hline \text{A, B, C, D, E, F} \\ \hline \text{label} : A \times B \times C \times D \times E \times F \rightarrow A \\ \text{content} : A \times B \times C \times D \times E \times F \rightarrow B \\ \text{temporal} : A \times B \times C \times D \times E \times F \rightarrow C \\ \text{causal} : A \times B \times C \times D \times E \times F \rightarrow D \\ \text{spatial} : A \times B \times C \times D \times E \times F \rightarrow E \\ \text{evaluation} : A \times B \times C \times D \times E \times F \rightarrow F \\ \hline \end{array}$$

where the first element describes the symbol that is used as the experience’s label, the second describes its content, the third, fourth, and fifth describe the temporal, causal, and spatial relations (if any), respectively, between the elements of the experience’s content, and the sixth describes the experience’s evaluation. Note that the evaluation is defined through a function that maps a symbol to an evaluation space. The symbol is intended to represent the evaluation criteria, such as “impact”, “usefulness”, “contentment”, “surprise”, and so on. 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 (?) are: anger, disgust, fear, surprise, sadness, and happiness.

Experiences should satisfy a number of properties. First, the domain and the range of the temporal, causal, and spatial relations are over the content of the experience. This is expressed accordingly.

$$\forall e : \text{CompoundExperience} \bullet \begin{cases} (\text{ran temporal}(e) \cup \text{dom temporal}(e)) \subseteq \text{content}(e) \\ (\text{ran causal}(e) \cup \text{dom causal}(e)) \subseteq \text{content}(e) \end{cases}$$

The third property states that if one sub-experience (or per- cept) leads to another, then the latter cannot temporally pre-cede the former. This is expressed accordingly.

$$\forall e_1, e_2 : \text{CompoundExperience} \bullet \begin{cases} (\text{label}(e_1) = \text{label}(e_2) \lor \text{content}(e_1) = \text{content}(e_2)) \Rightarrow e_1 = e_2 \end{cases}$$

The fourth property states that if two sub-experiences (or percepts) are spatially related, then they cannot be tempo- rally related.

$$\forall e : \text{CompoundExperience} ; e_1, e_2, : \text{Experience} ; s : \text{SpatialRelationships} \bullet (e_1, e_2, s) \in \text{spatial}(e) \Rightarrow ((e_1, e_2) \not\in \text{temporal}(e) \land (e_2, e_1) \not\in \text{temporal}(e))$$

Finally, we distinguish a special type of experience, a basic experience, in which all the contents of the experience are single percepts.

$$\text{BasicExperience} \equiv \begin{cases} e : \text{CompoundExperience} \bullet \langle \forall x : \text{content}(e) \bullet x \in \text{ran basic} \rangle \end{cases}$$

Manifesting Experiences We have defined what experiences are. But how are they manifested? We say agents have a capability to generate experiences by grouping existing experiences (whether they were percepts or compound experiences), labelling the resulting experience, and analysing it through the temporal $\prec$, causal $\rightarrow$, spatial $\triangleright$, and evaluation $\triangledown$ functions presented earlier. As such, we define the agent’s capability to manifest experiences accordingly.
where the first condition states that the resulting experiences did not exist in the old memory and that they are added to the new memory, and the second states that an experience is added to Memory only if its evaluation deems it worthwhile to remember (for instance, it is not neutral with respect to the evaluation space, or it was not expected by the agent). Different definitions for “worthwhile” may also be provided, such as the impact on belief update.

We remind the reader that experienced may also be recalled and modified (or even forgotten) over time via the recall function mentioned earlier.

**Actions as Experiences** We note that an experience can represent any combination of percepts and sub-experiences. If we flatten its content to obtain a set of single percepts, then these percepts may be of any type. They may be observed or imagined (where experiences with imagined percepts and their use are discussed in more detail in the following subsections). They may or may not contain agent, event, and action percepts. If the set did not contain any event or action percepts, then this implies that the experience describes a still image where nothing is happening. If the set contained only event percepts, then the experience describes natural events (such as rain, earthquakes, etc.). If the set contained action percepts, then the experience describes an event in which agents have acted.

Of particular interest to artificial intelligence is how to define single actions as experiences. The action percept describes a single event (such as driving) and the agent that carries out that event. However, actions usually have preconditions and post-conditions. As such, describing a single action with its pre and post conditions becomes important particularly for reasoning.

We say a basic action is defined as a basic experience, whose content contains only the percepts describing the preconditions of the action, the action percept describing the action itself, 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 accordingly.

\[
\begin{align*}
\forall e : \mathbb{P} \text{Experience} \mid m_1, m_2 : \text{Memory} \bullet \\
\text{experience}(m_1) & = (e \times m_2) \Rightarrow (e \subseteq m_1 \land e \subseteq m_2) \\
\wedge \\
\forall e : \mathbb{P} \text{Experience} \mid m_1, m_2 : \text{Memory} \bullet \\
\text{experience}(m_1) & = (e \times m_2) \Rightarrow \\
(\forall e' \in e \bullet \text{evaluation}(e') \notin \text{neutral}(\text{EvaluationSpace}) \lor \\
\text{evaluation}(e') \neq \text{expected}(\text{evaluation}(e'))) \\
\end{align*}
\]

A Note on Events We argue that natural events, such as rain, earthquakes, and so on, are not of much interest to agents. We believe this is because agents would only be interested in whether certain properties are satisfied or not in the environment, and this does not require distinguishing basic events. However, if cases arise in which basic events are needed, they are defined similarly to basic actions, where the only difference is that the action percept is replaced with a event percept.

**Beliefs**

Agents hold beliefs about itself, its environment, its feelings (if any), and so on. We distinguish between two different types of beliefs: factual beliefs and generic beliefs. Factual beliefs are beliefs that the agent beliefs are facts, while generic beliefs represent generic rules that the agent believes in. Factual beliefs may be divided further into two sub-categories: perceived beliefs and deduced beliefs. The former describes the perceived perceptions that one decides to believe. For instance, one may perceive the book on the shelf and decide to believe that the book is indeed on the shelve. We note that not all percepts may be believed by the agent. For example, one may be on the road on a hot summer day and perceive some water on the road. However, it may decide to believe that this is a mirage, and that there is no water on the road. Deduces beliefs on the other hand may be deduced based on past experiences and what one has learned from those experiences. For example, one may believe that it never rains when the sky is blue and that the sky is blue today. As such, one may have a deduced belief that it will not rain today. Deduces beliefs may also be the indirect result of an observed percept. For instance, Tristan may inform Mireille that Emily played the piano last night. Mireille may then belief that Tristan told her this piece of information, and this would be consider a factual belief since she actually perceived it herself and decided to believe it. Mireille may also decide to believe that Emily did play the
piano last night, and this would be considered a deduced belief since this is not something that she actually perceived herself.

As for generic beliefs, we say they represent rules that one believes hold in general. For example, “all men are mortal”, or “there exists birds that do not fly”. Generic beliefs are usually deduced by analysing one’s entire history of experiences and knowledge (or beliefs). They evolve with time as one acquires new knowledge. As such, they may or may not be true in the physical environment.

Belief is then defined accordingly.

\[ \text{Belief} ::= \begin{cases} \text{factual}(\text{Factual}) \\ \text{generic}(\text{Generic}) \end{cases} \]

Manifesting Beliefs In summary, we say that new beliefs are generated usually as a result of both new experiences and old beliefs. As such, we say the capability of generating new beliefs is defined through the function \( \text{believe} \), whose input is the agent’s memory and its output is the set of newly generated beliefs and the agent’s updated memory.

\[ \text{ManifestingBeliefs} \]
\[ \forall b : \text{Belief}; m_1, m_2 : \text{Memory} \quad \text{believe}(m_1) = (b \times m_2) \Rightarrow (b \not\subseteq m_1 \land b \subseteq m_2) \]

where the conditions state that the resulting new beliefs did not exist in the old memory and that they are added to the new memory. We remind the reader that beliefs are recalled and modified (or even forgotten) over time via the \textit{recall} function mentioned earlier. If an alternative function dedicate for belief update is needed then it may be added.

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. Desires are then defined accordingly.

\[ \text{Desire} \]
\[ \text{content}(\text{desire}) \subseteq \text{ran imagined} \]

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 shortly.

Manifesting Desires An agent generates new desires through its capability of desiring experiences.

\[ \text{ManifestingDesires} \]
\[ \forall d : \text{Desire}; m_1, m_2 : \text{Memory} \quad \text{desiring}(m_1) = (d \times m_2) \Rightarrow (d \not\subseteq m_1 \land d \subseteq m_2) \]

where the conditions state that the resulting new desires did not exist in the old memory and that they are added to the new memory. We remind the reader that desires are recalled and modified (or even forgotten) over time via the \textit{recall} function mentioned earlier, which may modify older desires.

Intentions

Intentions are desires with plans, where the actions in these plans are imagined ones that the agent has committed to realising, but has not realised yet. As such, before defining intentions, we first need to define plans.

However, we say a plan consists of a partial order of imagined occurrences which are either satisfied properties or agent actions. As such, we start by defining imagined satisfied properties (which we refer to as imagined experiences) and imagined agent actions.

An imagined experience is an experience that has not happened yet; i.e. the percepts constituting this experience are imagined.

\[ \text{ImaginedBasicExperience} == \{a \in \text{BasicExperience} \mid \text{content}(a) \subseteq (\text{ran imagined})\} \]

We note that when a plan’s definition includes properties that should be satisfied, these properties usually can be described through basic experiences, as opposed to compound experiences. If situations arise in which this is not the case, then one may think of using imagined compound experiences as opposed to imagined basic ones.

An action template is an imaginary action that has not happened yet. Again, all the percepts constituting this action are therefore imagined. We remind the reader that the content of an action consists of the pre-conditions, the post-conditions, and the action percept. Necessarily, the causal relation will define the fact that the action percept causes the post-conditions of the action.

\[ \text{ActionTemplate} == \{a \in \text{BasicAction} \mid \text{content}(a) \subseteq (\text{ran imagined})\} \]

An imagined occurrence is then defined accordingly.

\[ \text{ImaginedOccurrence} ::= \{\text{action}((\text{ActionTemplate})) \mid \text{prop}(\text{ImaginedBasicExperience})\} \]

A plan becomes a partial order of imagined occurrences.

\[ \text{Plan} == \big\langle \text{ImaginedOccurrence} \times \text{ImaginedOccurrence} \big\rangle \]

Finally, an intention is defined accordingly:

\[ \text{Intention} == \text{Desire} \times \text{Plan} \]

We note that when performing actions, the imagined action will be recorded as an experience with observed percepts, and this will introduce a new data structure to the memory.
Manifesting Intentions

The agent’s capability to generate new intentions is defined accordingly.

\[
\text{Manifesting\_Intentions} \quad \text{intend} : \text{Memory} \to (\mathbb{P} \\text{Intention} \times \text{Memory})
\]

\[
\forall i : \mathbb{P} \\text{Intention}; m_1, m_2 : \text{Memory} \quad \text{intending}(m_1) = (i \times m_2) \Rightarrow (i \not\subseteq m_1 \land i \subseteq m_2)
\]

where the conditions state that the resulting new intentions did not exist in the old memory and that they are added to the new memory. We remind the reader that intentions are recalled and modified (or even forgotten) over time via the recall function mentioned earlier, which may modify older intentions.

The Agent’s Definition

An agent is then a compilation of its own experiences (both percepts and compound experiences), beliefs, desires, and intentions; its capabilities to manifest (and recall) these formulas; its capability to perform other actions (such as making commitments, sending messages, walking towards a given destination, and so on), as well as its capability to represent its knowledge and reason over it. An agent is then defined accordingly.

<table>
<thead>
<tr>
<th>Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>name : Symbol</td>
</tr>
<tr>
<td>actionCapability : \mathbb{P} BasicAction</td>
</tr>
<tr>
<td>language : \mathbb{P} Symbol</td>
</tr>
<tr>
<td>KnowledgeRepresentationReasoning</td>
</tr>
<tr>
<td>percepts : \mathbb{P} Percept</td>
</tr>
<tr>
<td>ManifestingPerceptions</td>
</tr>
<tr>
<td>SpatialRelationships, EvaluationSpace</td>
</tr>
<tr>
<td>experiences : \mathbb{P} Experience</td>
</tr>
<tr>
<td>ManifestingExperiences</td>
</tr>
<tr>
<td>beliefs : \mathbb{P} Belief</td>
</tr>
<tr>
<td>ManifestingBeliefs</td>
</tr>
<tr>
<td>desires : \mathbb{P} Desire</td>
</tr>
<tr>
<td>ManifestingDesires</td>
</tr>
<tr>
<td>intentions : \mathbb{P} Intention</td>
</tr>
<tr>
<td>ManifestingIntentions</td>
</tr>
<tr>
<td>Recalling</td>
</tr>
<tr>
<td>SensoryMemory : \mathbb{P} Percept</td>
</tr>
<tr>
<td>WorkingMemory ::= exp({Experience}</td>
</tr>
<tr>
<td>LongTermMemory ::= exp({Experience}</td>
</tr>
<tr>
<td>Memory ::= sensory({SensoryMemory}</td>
</tr>
</tbody>
</table>

\[
\forall e \in \text{experiences}; b \in \text{beliefs}; d \in \text{desires}; i \in \text{intentions} \quad (e, b, d, i) \subseteq \text{LongTermMemory}
\]

\[
\forall x : \text{Memory} \quad x \subseteq \text{WorkingMemory} \Rightarrow x \subseteq \text{LongTermMemory}
\]

where actionCapability defines the actions that the agent is capable of performing, and the conditions above state that:

- first, the long term memory contains all experiences, beliefs, desires, and intentions; and second, the working memory is always a subset of the long term memory.

We remind the reader that not all agents will have all these capabilities, most usually 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 capable agent (based on the actionCapability), the sensor agent (based on the PerceivingCapability), the intentional agent (based on the capabilities resulting in manifesting experiences, beliefs, desires, and intentions), and so on. Although our research focuses is on the intentional agent, as illustrated by the remainder of this paper.

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 SocialWorld structure specified in the final schema of the final section and then detail a few properties of this logic.

Syntax

Here we define the syntax of X-BDI. We assume a set of symbols for agents, A, and for plans, P, are available. (Please note for the purposes of keeping our work within the page limited we are using a simplified version of plans that includes actions but not events. However, their inclusion would not affect the semantics of our logic.)

- If \( \varphi \) is a propositional Well Formed Formula then \( \varphi \in \text{X-BDI} \)
- If \( \varphi \in \text{X-BDI} \), \( \alpha \) is an agent, and \( p \) is a plan then \( E(\alpha, \varphi), B(\alpha, \varphi), D(\alpha, \varphi), G(\alpha, \varphi), I(\alpha, \varphi, p) \in \text{X-BDI} \)
- If \( \varphi \) is a propositional Well Formed Formula, \( \alpha \) is an agent, then \( \text{Action}(\alpha, \varphi) \in P \)
- If \( \varphi \in \text{X-BDI} \), \( G = \{ \alpha, \beta, \ldots \} \subseteq A \), and \( p \in \mathbb{2}^G \) then \( \text{CE}(G, \varphi), \text{JE}(G, \varphi), \text{SE}(G, \varphi, p), \text{CD}(G, \varphi), \text{JD}(G, \varphi), \text{SD}(G, \varphi, p), \text{CG}(G, \varphi), \text{JG}(G, \varphi), \text{SG}(G, \varphi, p), \text{CI}(G, \varphi, p), \text{JI}(G, \varphi, p), \text{SI}(G, \varphi, p) \in \text{X-BDI} \)
- If \( \varphi \) and \( \psi \in \text{X-BDI} \) then \( \neg \varphi, \varphi \lor \psi \in \text{X-BDI} \)

The meaning of the symbols are straightforward: B, D, and I are the classical ones, E and G stand for Experience and goal respectively. We use the letter C as a prefix to one of the previous symbols to mean Collective. Similarly, J stands for joint and S stands for Shared. Action has the obvious meaning. Thus, the expression:

\[
\text{SI}(\{\alpha, \beta\}, \text{WinBarcelona}, \{\text{Action}(\alpha, n), \text{Action}(\beta, \text{stadium})\})
\]

is a literal in X-BDI expressing that \( \alpha \) and \( \beta \) have the shared intention to see Barcelona play a football match by \( \alpha \) seeing the match on tv and \( \beta \) going to the stadium.
Semantics

We will use the SocialWorld structure as specified at the end of the previous section as the interpretation of the logic. Formally, we understand it 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 $\varphi$ is a percep resulting from an experience of $\alpha$ that is internalised by $\alpha$ as a belief then we can say that the formula $B(\alpha, \varphi)$ is true. Similarly, we can define the semantics of formulae like $D(\alpha, \varphi)$ or $G(\alpha, \varphi)$. As argued before, agents have perceps about other agents, more concretely about the desires, goals, and beliefs of other agents. This means that we will give semantic to nested expressions by looking into the perceptive memory of the agents. For instance, $\alpha$ may have seen an agent $\beta$ in the environment watching a football match and jumping with joy when Barcelona scores a goal and thus $\alpha$'s desire that Barcelona wins, and that can be represented in the logic as $B(\alpha, D(\beta, \text{WinBarcelona}))$. In this sense the $D(\beta, \text{WinBarcelona})$ inside the previous formula has to be understood as part of a percept of $\alpha$ and not as a percept itself. That is, $\beta$ may be pretending to like Barcelona. Differently from classical Kripke approaches we assume that the environment and the mental states of agents evolve due to events and agent actions and that we cannot know a priori how the next state will be computed. Thus, we assume that at any moment we can have the ‘current’ SocialWorld to interpret formulae of the logic but we cannot reason across actions of the agents. In this way semantics do not have the equivalent of the accessibility relationships usual in Dynamic Logics. In particular there is no way we can know how an action modifies the state of an agent.

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

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. The satisfaction of a plan consists of the satisfaction of each individual action in the plan.

\[
\langle W, A \rangle \models \text{Action}(\alpha, \varphi) \text{ iff } \exists \beta \in A. (\alpha, \varphi) \in \beta.\text{actions}
\]

\[
\langle W, A \rangle \models p \text{ iff } p \in 2^\varphi \text{ and } \forall a \in p. \langle W, A \rangle \models a
\]

Experiences: A model satisfies that $\alpha$ had an experience $\varphi$ if it has a memory of it in its repository of experiences (as defined by the variable experiences in the MentalAgent schema in Section 2). A group of agents had a common experience tagged as $\lambda$ if all had an experience that semantically entails $\lambda$. Similarly, a group of agents had a joint experience tagged as $\lambda$ if all had an experience that semantically entails $\lambda$ and they recognise each other as part of the experience. Finally, a group of agents had a shared experience tagged as $\lambda$ if it was a joint experience and furthermore they all actively participated in it.2

\[
\langle W, A \rangle \models E(\alpha, \varphi) \text{ iff } \varphi \in \alpha.\text{experiences}
\]

\[
\langle W, A \rangle \models CE(G, \lambda) \text{ iff }
\forall \alpha \in G. \exists \varphi \in X-BDI
\langle W, A \rangle \models E(\alpha, \varphi) \land \varphi \rightarrow \lambda
\]

\[
\langle W, A \rangle \models JE(G, \lambda) \text{ iff }
\forall \alpha, \beta \in G. \exists \varphi \in X-BDI
\langle W, A \rangle \models E(\alpha, \varphi) \land \varphi \rightarrow \lambda \land \varphi \rightarrow \beta
\]

\[
\langle W, A \rangle \models SE(G, \lambda, p) \text{ iff }
\langle W, A \rangle \models JE(G, \lambda) \text{ and }
\forall \alpha \in G. \exists \text{Action}(\alpha, \varphi) \in p.
\langle W, A \rangle \models \text{Action}(\alpha, \varphi) \land \varphi \rightarrow \lambda
\]

Beliefs: The semantics for beliefs are 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.

\[
\langle W, A \rangle \models B(\alpha, \varphi) \text{ iff } \varphi \in \alpha.\text{beliefs}
\]

\[
\langle W, A \rangle \models CB(G, \varphi) \text{ iff }
\forall \alpha \in G. \langle W, A \rangle \models B(\alpha, \varphi)
\]

\[
\langle W, A \rangle \models JB(G, \varphi) \text{ iff }
\forall \alpha, \beta \in G. \langle W, A \rangle \models B(\alpha, \varphi) \land B(\beta, B(\alpha, \varphi))
\]

\[
\langle W, A \rangle \models SB(G, \varphi, p) \text{ iff }
\langle W, A \rangle \models JB(G, \varphi) \text{ and }
\exists \lambda \in X-BDI. \langle W, A \rangle \models SE(G, \lambda, p) \land \lambda \rightarrow \varphi
\]

The same approach used for beliefs can be followed to Desires and Goals. Next we give semantics to 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.

\[
\langle W, A \rangle \models I(\alpha, \varphi, p) \text{ iff } (\varphi, p) \in \alpha.\text{intentions}
\]

\[
\langle W, A \rangle \models CI(G, \varphi, p) \text{ iff }
\forall \alpha \in G. \langle W, A \rangle \models I(\alpha, \varphi, p)
\]

\[2\text{We simplify matters by abusing notation and writing } a \in p \text{ to mean that action } a \text{ is 'part' of plan } p.\]
Motivating the interest in a shared experience. Agents may have different motives for engaging in shared experiences, based on their own personal beliefs, desires, and goals. However, when one agent realises that it would like to fulfill 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 commit to the goal of realising the shared experience. In multiagent systems, this could be achieved through argumentation. This step describes the agent’s investigation of whether a joint goal for achieving the shared experience may be realised.

Investigating a joint goal. If agents agree that they are all interested in realising a shared experience, then a joint goal to achieve the shared experience is born (i.e. it becomes true).

Realising a joint goal. If agents agree 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

$D(c, SE\{\{c, anyone\}, g, inanyway\})$ (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 anyone). This could be either because he is dependent on her, or possibly secretly in love with her. As such, Bill’s belief base contains the following:

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

1. Investigating a joint goal. 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 (j) 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:

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

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

And deducing that:

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

Similarly, Bill modifies his belief base to contain the following:

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

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

And deducing that:

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

As such, the following now holds:

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

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.
2. Realising a joint goal. Different protocols for agreeing on commitments may exist in multiagent systems. However, for this simple scenario, we assume that if one asks another if they are interested to join them in a commitment, there then constitutes a commitment to realise $x$. And given that they are communicating their commitments, then this implies that they are also aware of each other’s commitments. As such, Carla’s belief base is updated to contain the following:

\[ G(c, SE(\{c, b\}, g, inanyway)) \]  
(k)
\[ B(c, G(b, SE(\{c, b\}, g, inanyway))) \]  
(l)
And the following is deduced:

\[ B(c, JG(\{c, b\}, SE(\{c, b\}, g, inanyway))) \]
Similarly, Bill’s belief base is updated to contain the following:

\[ G(b, SE(\{c, b\}, g, inanyway)) \]  
(m)
\[ B(b, G(c, SE(\{c, b\}, g, inanyway))) \]  
(n)
And the following is deduced:

\[ B(b, JG(\{c, b\}, SE(\{c, b\}, g, inanyway))) \]
As such, the following now holds:

\[ JG(\{c, b\}, SE(\{c, b\}, g, inanyway)) \]  
(o)
Note that (o) is deduced from the axiom (5), the belief literals (k) and (m).

3. Investigating a shared intention. After both agents agree that they share the goal 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.

4. Realising a shared intention. 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:

\[ I(c, SE(\{c, b\}, g, p), p_c) \]
\[ B(c, I(b, SE(\{c, b\}, g, p), p_b)) \]
\[ B(c, p_c \neq \emptyset \land p_b \neq \emptyset) \]
And the following is then deduced:

\[ B(c, JI(\{c, b\}, SE(\{c, b\}, g, p), p)) \]
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 to contain the following:

\[ I(b, SE(\{c, b\}, g, p), p_b) \]
\[ B(b, I(c, SE(\{c, b\}, g, p), p_c)) \]
\[ B(b, p_c \neq \emptyset \land p_b \neq \emptyset) \]
And the following is deduced:

\[ B(b, JI(\{c, b\}, SE(\{c, b\}, g, p), p)) \]
As such, the following now holds:

\[ JI(\{c, b\}, SE(\{c, b\}, g, p), p) \]

5. 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 to contain the following:

\[ B(b, SE(\{c, b\}, g, p)) \]
And the shared experience is said to have been realised:

\[ SE(\{c, b\}, g, p) \]

Background

Individual actions are dictated by individual intentions, which are the result of individual beliefs, desires, and goals. But how do collective actions come about? In this section, we relate our work to that in the field of Philosophy (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 (Gilbert 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 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, goals, 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
For that, the are not used to describe the groups’ intentions as a whole.

ties, we note that these collective modalities are simply used and a group of friends who decide to jog together. there is something different between this group of people 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. For example, 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 achieving a joint goal. However, before we dwell on the shared modalities, we first introduce the philosophers’ different views, and then compare them to our proposed model.

Searle (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 two couples 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 fulfill 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.

However, Searle has been strongly criticized 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 (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).

Our model is similar to Bratman’s in the sense that an intention is shared only if there is a shared plan of action that meshes and coordinates the individual agents’ 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. Also, 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.

However, 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 (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.

Conclusion

In this paper we have proposed the notion of shared experience as a fundamental construct to underpin collective intentionality. We have provided a formal account of agency in a standard formal specification language based on the perceptions derived from the observation of the environment, the agents and their actions. We have also built a logical formalism to describe agents reasoning whose semantics is based on this agent model. 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 from the Z specification and the X-BDI reasoning system. We are using the theory outlined in this paper to build social browsers that will enable human users, supported by software agents, to share online cultural experiences such as jointly visiting a museum or gallery from their homes.

References


