Towards a Computational Ontology of Mind

Roberta FERRARIO[1] and Alessandro OLTRAMARI[1,2]
[1]Laboratorio di Ontologia Applicata, ISTC-CNR
[2]Dipartimento di Scienze della Cognizione e della Formazione, Università di Trento
Via Solteri 38
38100 Trento (Italy)
{ferrario; oltramari}@loa-cnr.it}
URL: http://www.loa-cnr.it

Abstract. The main goal of this paper is a preliminary characterization of the categories of the mental, able to fit and integrate the foundational ontology DOLCE (*a Descriptive Ontology for Linguistic and Cognitive Engineering*); we will call this core ontology COM (*Computational Ontology of Mind*).
The idea of COM emerges from the need of a conceptual clarification from the standpoint of formal ontology of the entities that play a role in agent technologies for information systems.
Based on philosophical tradition, we have singled out a central relation in the realm of the mental: *aboutness*. In our proposal *aboutness* connects a mental state with a mental object, at a certain time, and with respect to a given intentional agent.
Thus, in the paper we will give a first analysis of these entities, mainly focused on mental objects and their characteristics. We are also specifying the basic features of mental states and intentional agents, exploiting ontological categories and relations implemented in DOLCE.

"We exist, as the thinking things we are, only thanks to a baffling dance of brains, bodies, cultural and technological scaffolding" [1].

Introduction

Agents are not only a technological resource, but also a metaphor used in the design of complex systems, a model for the study of the actual world and, most of all, a ground where many disciplines meet, as shown by their multifaceted applications. Electronic commerce, bioinformatics, computational biology, Semantic Web, Grid Computing are only some of the many fields in which agents have begun to be really widely used. On the other hand, the paradigm of agents is based on concepts imported from many disciplines: the belief-desire-intention modeling from philosophy, interaction and dialogue, trust and delegation from social sciences, ecosystems from biology, and tools developed in many areas, like temporal, modal and epistemic logics, and game theory, just to quote some examples.
One of the most important challenges that the agents community is facing is that of building "open multi-agent systems spanning multiple application domains, and involving heterogeneous participants developed by diverse design teams" [2]. If, as envisaged, this stage will be reached, then such agents will embed different conceptualizations of the common "environment" where they are interacting. At this point they will need a tool that play the role of a bridge between the different conceptualizations. This is what foundational ontologies, such as DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*) are supposed to supply. They are characterized by a very rich formal language and by the presence of many relations among categories and one of their main aims is that of characterizing the intended meaning of terms through different communities of use, in

order to make agents (even heterogeneous ones) capable of communicate with each other and share information about the environment. Obviously, other agents are also part of the environment, and thus agents need to be able to represent them and communicate about them. From this follows that a foundational ontology ought to include a module dedicated to intentional agents and their mental features.
The main goal of this paper is that of sketching a preliminary characterization of the categories of the mental, able to fit (or possibly to amend) and integrate the foundational ontology DOLCE [3]; we will call this module COM (*Computational Ontology of Mind*).
The idea of COM emerges from the need of a conceptual clarification from the standpoint of formal ontology of the entities that play a role in agent technologies for information systems, like mental attitudes, objects, processes, etc.
The paper focuses on the internal dimension, disregarding for the moment the external dimension of the interaction with other agents and the relation between internal states and what originates them. Namely, the problem of how entities existing outside the agent can cause mental phenomena will be left unanalyzed for now.
As happens with other domains, there are many possible characterizations of the realm of the mental; a long term goal is thus that of building several modules representing some relevant or particularly meaningful positions and of integrating these modules via formal relations.
The starting point of COM is the sketch of a representational approach, inspired by J. A. Fodor's "Computational Representation Theory of Mind" (CRTM) [4] from the philosophical side and by the BDI (Belief, Desire, Intention) approach from the computer science side. In the paper we won't concentrate on questions like whether it is legitimate to compare some internal state of a machine with a human belief or intention; we will adopt what Daniel Dennett [5] has called the "intentional stance", i.e. we will use mental entities as means to describe and predict the behaviour of machines. Sometimes, even if an explanation based on the physical functioning of a machine is available, it is easier, at a descriptive level, to give a representation in intentional terms, as Michael Wooldridge pointed out in [6].
Both CRTM and BDI have been particularly influential in their own discipline and they are particularly fit to integrate a computational perspective (in particular, BDI agents have successfully been implemented in several applications); these are the main reasons why we have decided to choose them as starting point.

1 Intentional Agents

In this section we want to informally enucleate the main features of what is meant by the locution "intentional agent" in the literature in philosophy and in artificial intelligence, in order to eventually characterize it within the DOLCE ontology. This characterization will abstract away from the physical appearance in which the intentional agent is embodied. This means that the representation of intentional agent should approximate a human being, a robot or a software program.
In [7] Dennett distinguishes between what he calls "first-order intentional systems" and "second-order intentional systems", where the first-order are those systems that can be in some mental states, like belief, desire, intention, but cannot entertain a mental state about another mental state, like a belief, desire, or intention about another belief, desire or intention. Even though we are interested in complex intentional agents, capable of displaying a "second-order behaviour", in this paper we will only analyze them while behaving in the "first-order manner".
In [8] John Pollock defines an agent as "any system capable of acting on its environment to render it more congenial to its continued survival".

Starting from this, in philosophy and in artificial intelligence an intentional agent has often been defined as one that always acts on the basis of its desires, beliefs and intentions and possesses the ability to adapt to the environment, updating and revising these desires, beliefs, and intentions.
A more thorough formulation is known in artificial intelligence as the BDI paradigm (Belief, Desire, Intention) [9]. The success of this model is only in part due to the usefulness shown by the applications using it; a very important reason is for sure the solid philosophical model that is at its bases, inspired by the work of Michael Bratman [10].
As outlined by Michael Wooldridge in [11], there are other alternative paradigms in the AI literature, such as logical architectures, reactive architectures (as those described by Rodney Brooks in [12], [13] and [14]) and hybrid agents of the stratified architectures. Nonetheless, we have decided to focus on the BDI approach as, with its representational character, is – as far as we know – the closest to the common sense way of representing mind and its processes.
The BDI model is based on three fundamental components:

- *Beliefs*: they represent the knowledge of the agent on the world;
- *Desires*: they represent the states the agent wants to reach, so, in a derived sense, its goals;
- *Intentions*: they represent the desires that the agent is committed to achieve.

In particular, intentions play a very important role, as they constrain the actions and the behaviour of agents, they determine whether the agent keeps pursuing a goal or quits it and embraces another one. This element of commitment is very important, as agents are resource-bounded and they cannot pursue each single desire they have, they ought rather to focus on some of them and neglect others. Philip Cohen and Hector Levesque in [15] claim that an intention is abandoned either when it has been achieved, or when the beliefs of the agent change and the goal seems to have become unreachable. Another circumstance we could add is when the beliefs have changed and another goal seems preferable. Here both perception and reasoning play important roles.
In the BDI literature a lot of attention is dedicated to the capability that agents must possess of decomposing plans in sub-plans and thus goals in sub-goals. This determines some sort of compositionality of goals.
When it acts, an agent commits to an activity that is directed toward a goal that the agent has embraced on the basis of considerations on the available options. Moreover, the agent is aware both of the fact of being committed, and of the fact that the action is aimed at a certain goal.
Another very interesting characterization of agents we want to rely on is the one given by Cristiano Castelfranchi:

> "At a very basic level, an agent is any entity able to act, i.e., to produce some causal effect and some change in its environment".
>
> […]
>
> "In other terms, the agent's behaviour is aimed at producing some result: thus we are talking of a *goal-oriented* action and of a goal-oriented agent. Among goal-oriented systems I will consider in particular *goal-directed* systems. In these systems not only action is based on perception, but the latter is also the perception of the action's effects and results, and the agent regulates and controls its actions on such a basis. *The agent is endowed with goals*, i.e., internal anticipatory and regulatory representations of action results" [16].

So, here we find a more fine-grained differentiation among agents with goals, namely that between goal-oriented and goal-directed agents. The notion of intentional agent we want to

characterize and use is closer to the goal-directed side. An agent is goal-directed in two senses: more generally, because its action ends with the achievement of the goal and, more specifically, because it builds a mental representation of the goal, of the action necessary to its achievement, and of the resulting consequences.
Another very central feature of intentional agents is the fact that they can be (and usually are) located in a social environment and thus they must possess social capabilities or, in other words, they must be able to interact. With interaction what is usually meant is that intentional agents can be influenced by other agents in their mental activities, in the achievement of their goals, or in the execution of their tasks. This interaction can be either direct or mediated by the environment. In order to talk about interaction, it is necessary to presuppose some sort of communication among agents, not necessarily verbal.
So far, we have reasoned on the main features that are described in the literature on intentional agents. Now we have to understand which are the requirements an entity must satisfy in order to be categorized as an "intentional agent".
In [6] Wooldridge gives a set of properties that an intentional (or rational, in Wooldridge's terms) agent should possess: *autonomy*, *reactivity*, *proactiveness, social ability*.

Autonomy: this property, for artificial agents, normally means that the agent is not directly influenced in its decisions by the designer, or, more generally, by other agents.

Reactivity: an intentional agent must be able to properly react to the stimuli coming from the external environment; that is to say, every new perception must determine a revision of its mental states (even to keep them unaltered, it needs to reconsider them).

Proactiveness: the agent must not only be able to react to the environment, it must also be capable to take the initiative, thus showing goal-directedness, in the sense explained by Castelfranchi.

Social Ability: the intentional agent must be able to interact with other agents. Some of the external stimuli are originated by other agents, so even these can be source of new mental states.

In the next section we are going to introduce the theoretical context in which intentional agents possibly fit in, also proposing a formal characterization of relevant mental entities.

2 Mental Entities

2.1 The General Framework

According to classical argumentations in philosophy (see [17], [18], [19]), mental attitudes like "The Pope's *belief* that God exists", "Eddy's *fear*[1] of the dark" or "Neil Armstrong's *desire* to bring the craft down safely in the Sea of Tranquillity" correspond to those peculiar occurrences called states. In this respect, mental attitudes differ from mental *acts* ([19], p. 39), namely dynamic occurrences like "Tom's *watching* the football match on TV", "John's *noticing* a strange scar, like a bite, in Mina's neck", etc. In DOLCE this distinction between static and dynamic occurrences has been modeled on the basis of two general properties:

[1] Even if not considered in the BDI literature, fears are treated as mental attitudes in the philosophy of mind.

cumulativity and *homeomericity*[2]. Intuitively, we may say that *states* are *cumulative* since the mereological sum of two instances of the same occurrence-type maintains the same occurrence-type (e.g the sum of two instances of "desire" is still of the type "desire"). Moreover, states are *homeomeric* since all their temporal parts are described by the very same expression used for the whole occurrence (e.g. every temporal part of "Bush's desire to be re-elected" is caught by the very same expression). Therefore we are going to refer here to mental attitudes as kinds of states. On the nature of mental attitudes there are different positions: some authors, like Dennett [5] and Churchland [20], represent them as propositions; others, like Crane [19], distinguish propositional attitudes from perceptions, and so on and so forth. In our proposal, we are simply considering mental attitudes to be states related to "mental objects": traditionally, this relation has been called "aboutness"[3] [19]. There are two possible ways of representing "aboutness": either it can directly connect the mental state of an agent with something which is outside its mind, or it can connect it with something that is internal to the agent. These two options gave rise to the core philosophical dispute between internalists and externalists: in general, externalists accept both possibilities, claiming that some mental states are directly connected to the outside world and admitting that others refer to mental objects; internalists, on the contrary, deny any direct connection to the external world. Although focused on the internal direction of aboutness, our position here is generically externalist: we simply choose to investigate the structure "mental state-aboutness-mental object" in order to give an account of the inner states of intentional agents, postponing the exploration of their *public life* to further work (see section 3). Like in Fodor's Computational Representational Theory of Mind (CRTM), we don't care about the intrinsic distinction between the conceptual or non-conceptual nature of mental objects: we consider mental objects to be nothing but representations of something else, both in the case they pertain to perceptual states and mental attitudes. We are not going to argue here about the mechanisms underlying those complex phenomena like perceptions: we are just adopting a modeling strategy to conform mental objects.

From a formal-ontological standpoint, the category of mental object has been characterized in DOLCE as a subtype of the class of non-physical objects:

"non-physical objects are divided into *Social Objects* and *Mental Objects* according to whether or not they are dependent on a community of agents. A private experience, for instance, is an example of *mental object*" ([3], p.17).

In conclusion, let us preliminarily say that in COM aboutness connects, at a certain time, a mental state with a mental object, with respect to an intentional agent. In other words, at a certain time, an intentional agent has a mental state about a mental object.

2.2 Formal Characterization

In this section we are going to formally characterize and specify which types of mental objects can be singled out, which are the mutual relations among them, with mental states and time[4].

---

[2] See [3], pp. 17-18; p. 21; p. 28. In this paper we focus only on *states*, avoiding to specify all the distinctions within other kinds of occurrences (processes, achievements and accomplishments) with respect to the realm of the mental.

[3] Here we are only referring to the domain of intentionality. We can obviously have a book about a subject, an information about a concert etc., but we are restricting our attention to this specific domain.

[4] See fig.1 for the integrated taxonomy of DOLCE and COM.

First, we mostly informally introduce the predicates of DOLCE we need in the theory (see [3] for the axiomatization and more details). Notice that DOLCE is based on a fundamental distinction between *endurants* and *perdurants*, the former corresponding to entities that *are in time* (tables, cars, trees, Stephen Hawking, my idea of freedom, etc.) the latter corresponding to entities that *happen in time* (conversations, wars, dives, concerts, weddings, etc.). Very roughly, endurants point to objects and perdurants point to events.
We import from DOLCE the following categories and relations:

- PRE(x, t) stands for "*x being present at time t*";
- PC(x, y, t) stands for "*the endurant x participates to the perdurant y during the time t*";
- OSD(x, y) stands for "*x is one-sided specific constant dependent on y*" (this is exemplified by the specific dependence relation holding between a mental object and an intentional agent);
- $\mathsf{ql}_{\mathsf{T}}(t, x)$ stands for "*t is the temporal extension of x*", where x can be an endurant, a perdurant or a quality;
- *APO*(x) stands for "*x is an agentive physical object*" (this category captures the notion of intentional agent we are dealing with in the paper);
- *MOB*(x) stands for "*x is a mental object*", namely a peculiar non-physical endurant dependent on an intentional agent;
- *ST*(x) stands for "*x is a state*", namely a perdurant for which cumulativity and homeomericity hold.

Three different relations of dependence are introduced in COM.

(D1) $\mathsf{GED}(x, y) \equiv_{df} \Box(\exists t(\mathsf{PRE}(x, t)) \rightarrow \exists t'(\mathsf{PRE}(y, t')))$[5] (*General Dependence*)
(D2) $\mathsf{HD}(x, y) \equiv_{df} \Box(\exists t(\mathsf{PRE}(x, t)) \rightarrow \forall t(\mathsf{PRE}(x, t) \rightarrow \exists t'(t' < t \wedge \mathsf{PRE}(y, t'))))$[6] (*Historical Dependence*)
(D3) $\mathsf{DHD}(x, y) \equiv_{df} \mathsf{HD}(x, y) \wedge \neg\exists z\,(\mathsf{HD}(x, z) \wedge \mathsf{HD}(z, y))$ (*Direct Historical Dependence*)

We define *General Dependence* (D1) as the weakest among the dependence relations, since it does not commit on any temporal relations of precedence, coincidence or consequence between x and y; on the contrary *Historical Dependence* (D2) defines a relation of precedence of y with respect to x. *Direct Historical Dependence* (D3) is a restriction of (D2), excluding the cases in which there can be a z that is intermediate between x and y with respect to the relation of *Historical Dependence*.

Two basic types of mental objects can be defined in COM, *percepts* (*PRC*) and *computed objects* (*COJ*), being the former independent on any other mental object and the latter historically dependent on at least another mental object:

(D4) $PRC(x) \equiv_{df} MOB(x) \wedge \neg\exists y(MOB(y) \wedge \mathsf{GED}(x, y))$ (*Percept*)
(D5) $COJ(x) \equiv_{df} MOB(x) \wedge \exists y(MOB(y) \wedge \mathsf{HD}(x, y))$ (*Computed Object*)

x is a mental object if and only if it is a *percept* or a *computed object*:

---

[5] In this module we shall adopt the simplest quantified modal logic, namely S5 plus the Barcan Formula [21]. In addition, note that in DOLCE every non abstract entity is present in at least a possible world.

[6] In COM we are going to use the relations of temporal precedence "<" and "Starts" defined in Allen and Hayes' theory of time [22], that formalizes a discrete and linear time using the binary primitive *meets* (||) between convex and extended intervals.

(A1) $MOB(x) \leftrightarrow (PRC(x) \vee COJ(x))$

In order to reflect the "private" nature of mental objects, we want to exclude that the same mental object can depend on two different intentional agents:

(A2) $(MOB(z) \wedge APO(x) \wedge APO(y) \wedge \mathsf{OSD}(z, x) \wedge \mathsf{OSD}(z, y)) \rightarrow x{=}y$

Here we show two examples of the above distinctions[7]:

1) Dante sees (Beatrice)
2) Dante believes (to have seen (Beatrice))
3) Dante remembers[8] (to have believed (to have seen (Beatrice)))

In 1) we have a *percept*, while in 2) and 3) we have *computed objects* spread in different time intervals.
A *percept* is connected to a dimension that is external with respect to the intentional agent: we say that it represents *entities* of the world[9]. We chose the general label *computed* in order to stress that *mental objects* belonging to this category are indirect, namely they are the result of the computational processes that occur every time that an input (external or internal) is processed. Within the category of *computed object* we distinguish *computed belief* (*COB*)*, computed desire* (*COD*) and *computed intention* (*COI*).

(D6) $COB(x) \equiv_{df} COJ(x) \wedge \exists y(PRC(y) \wedge \mathsf{DHD}(x, y))$ (*Computed Belief*)
(D7) $COD(x) \equiv_{df} COJ(x) \wedge \neg\exists y(PRC(y) \wedge \mathsf{DHD}(x, y)) \wedge \exists z(COB(z) \wedge \mathsf{DHD}(x, z))$ (*Computed Desire*)
(D8) $COI(x) \equiv_{df} COJ(x) \wedge \neg\exists y(PRC(y) \wedge \mathsf{DHD}(x, y)) \wedge \neg\exists z(COB(z) \wedge \mathsf{DHD}(x, z)) \wedge \exists w(COD(w) \wedge \mathsf{DHD}(x, w))$ (*Computed Intention*)

(D6)-(D8) establish a "nesting" within the distinct kinds of *computed object*. Before giving some examples of this nesting we introduce the relation of *aboutness* and the concept of *mental state*.

(D9) $\mathsf{AB}(x, y, z, t) \equiv_{df} APO(x) \wedge ST(y) \wedge MOB(z) \wedge \mathsf{OSD}(z, x) \wedge \mathsf{PC}(x, y, t) \wedge \mathsf{PC}(z, y, t)$ (*Aboutness*)

What (D9) says is that a state y of the intentional agent x is about a mental object z at time t if and only if the agent x participates to the state y at t and the mental object z also participates to y at t, being z one-sided specific constant dependent on x.
Given this relation of *aboutness*, holding between a state y of an intentional agent x and a mental object z at a time t, we can define four kinds of mental states, reflecting the partition of mental objects:

(D10) $PES(y) \equiv_{df} \exists x, z, t(\mathsf{AB}(x, y, z, t) \wedge PRC(z))$ (*Perceptual State*)
(D11) $BEL(y) \equiv_{df} \exists x, z, t(\mathsf{AB}(x, y, z, t) \wedge COB(z))$ (*Belief*)

---

[7] We use the parenthesis to identify Dante's mental objects. In these and in the following examples we are taking a third-person perspective: as God's Eye, we are supposing to watch inside other minds.

[8] We consider remembrances as mental attitudes too.

[9] We don't commit on the characteristics of this relation of "representation", nor on the way in which *percepts* and *computed objects* result from the interaction of the intentional agent with the world. In general, the supporters of CRTM talk about suitable *transducers* transforming input from the environment in order to let them be processed by the intentional agent.

(D12) $DES(y) \equiv_{df} \exists x, z, t(\mathsf{AB}(x, y, z, t) \wedge COD(z))$ (*Desire*)
(D13) $INT(y) \equiv_{df} \exists x, z, t(\mathsf{AB}(x, y, z, t) \wedge COI(z))$ (*Intention*)

(D14) $MES(x) \equiv_{df} PER(x) \vee BEL(x) \vee DES(x) \vee INT(x)$ (*Mental State*)
(D15) $MAT(x) \equiv_{df} BEL(x) \vee DES(x) \vee INT(x)$ (*Mental Attitude*)

Therefore, *mental states* are connected each other exploiting the formal relation of *direct historical dependence* holding among the *mental objects* related to them. Let's consider three examples.

i) "Courtney believes to have heard a shot coming from Kurt's room"
ii) "Homer wants to drink a cold Duff beer"
iii) "Neo intends to swallow the red pill"

In i) Courtney's *belief* is about a *computed belief* direct historical dependent on a *percept* of a certain sound.
In ii), Homer's mental attitude of *desire* is about the *computed desire* of drinking a Duff cold beer, that is direct historical dependent on a *computed belief*, for example that Duff beer is more refreshing than a hot cup of tea. Moreover there must have been a corresponding mental attitude of *belief* about that *computed belief*.
iii) has the most complex structure. Neo's mental attitude of *intention* is about the *computed intention* of swallowing the red pill, being this mental object direct historical dependent on at least a *computed desire*, for example Neo's desire of following Morpheus and Trinity, namely a *computed desire* direct historical dependent on at least a *computed belief*, e.g. Neo's belief that the world is just an illusion. Finally, this *computed belief* would be direct historical dependent on a *percept*, corresponding in Neo's mind to the discovery that "the world is just an illusion produced by the Matrix".
The resulting picture fits the main principles of the BDI approach we showed in section 1: when an agent takes a decision, its internal chain of states follows a rationale. In particular, as we have already mentioned in the first section of the paper, desires can be seen as *goals* an intentional agent has to reach by suitable strategies constrained to its knowledge of the world.
In our proposal *computed beliefs* are different from other *mental objects* because the existence of *percepts*, *computed desires* and *computed intentions* is bind to the duration of the *mental states* that are about them, while a *computed belief* requires at least that a *mental attitude* is present at the beginning of its life. It is then able to "survive", stored in the memory of the intentional agent. This distinction captures the general notion of the *Tacit Belief Box* [5], which is a sort of knowledge repository the intentional agent can exploit in order to form desires and intentions.

(A3) $COB(z) \rightarrow \exists x, y, t, t'(\mathsf{AB}(x, y, z, t) \wedge \mathsf{ql}_{\mathsf{T}}(t', z) \wedge (\mathsf{Starts}(t, t') \vee t = t'))$[10]
(*Computational Survival*)
(A4) $(PRC(z) \vee COD(z) \vee COI(z)) \rightarrow \exists x, y, t(\mathsf{AB}(x, y, z, t) \wedge \mathsf{ql}_{\mathsf{T}}(t, z))$
(*Computational Duration*)

But what happens when more than one agent is involved in a system? This is a tentative picture exploiting COM:

4) Neo fears (that he is the One)

[10] As in DOLCE, here we do not consider any specific theory of time, therefore *Starts* has to be intended as in the sense specified in [22].

5) Trinity believes (that Neo fears (that he is the One))
6) Morpheus suspects (that Trinity believes (that Neo fears (that he is the One)))

Here the temporal dimension does not seem relevant: even if 4), 5) and 6) occur at the same time, we will nevertheless have distinct mental states, since we are considering three different intentional agents with their own private states. In 5) Trinity forms a *computed belief* of the internal state Neo is in. So does Morpheus with respect to Trinity in 6). At different levels of complexity, Trinity and Morpheus, observing the behaviour of Neo (Trinity), and the behaviour of Trinity (Morpheus) try to represent what is happening in someone else's thoughts: nevertheless, they are dealing with *computed objects* deriving from *percepts* (namely, from the observation of each other's behaviour). Anyway this is only a partial and rough sketch of the further problems we have to work out in order to extend COM.

2.3 Intentional agents reloaded

In section 1, following the literature, we have given some informal specifications of the features we want an intentional agent to possess. Now, after the formal characterization of mental objects and mental states in DOLCE, in this section we preliminarily explore some consequences on the relations between intentional agents and their mental objects and states.
From what stated above, we know that when intentional agents are in a certain mental state, they participate to such mental state together with the mental object this state is about. But does an agent need to be constantly entertaining some mental state in order to be considered intentional?
Exploiting (A3) and (A4) we could assess a simple requirement: an agent, in order to be intentional, must have stored (for instance, in its knowledge base if it is a software agent) a mental object, more precisely, a computed belief. Intuitively, the fact of having a computed belief ensures that the agent, at some time has been in a state of belief about this computed belief and thus, it is capable of being in such a state.

So, formally we could introduce this axiom:

(A5) $(APO(x) \wedge \mathsf{PRE}(x, t)) \rightarrow \exists z(COB(z) \wedge \mathsf{PRE}(z, t) \wedge \mathsf{OSD}(z, x))$
(*Intentional Condition*)

The meaning of the axiom is the following: if x is an intentional agent which is present at the time t, there is at least a *computed belief* z that is present at t and one-sided constant dependent on x, even though no mental state of the intentional agent is about it. So, in order for an agent to be intentional, it must have some computed belief, even though it could be not aware of having such a belief (for instance, because it is momentarily "asleep").
We are aware of the fact that this doesn't explain the case of agents with a disposition to believe but that have not believed anything yet. From a philosophical standpoint, this is undoubtedly problematic and we certainly have to investigate this aspect more thoroughly, but for the moment, for the sake of simplicity, let's say that these agents have "innate" beliefs, in the case of human agents, and established by the designer in the case of artificial agents.
Far from claiming that our present model can exhaust the subject, we think that the road we are running along can bring us to a comprehensive model of relevant cognitive dimensions of intentional agents.

3 Conclusions and Future Issues

Agents' technologies are becoming a more and more widespread paradigm in various computer science applications. As happens with humans, in order to be more effective in their interactions, they need to be able to reason on each other's mental dispositions. For this reason, an ontology embedding a cognitive agent's vision of the world cannot lack a core ontology describing the realm of the mental. The aim of this paper was that of providing a first sketch of such sub-ontology for DOLCE. In the sequel of this work still there are many directions that can and ought to be deepened; first of all, we need to give a more precise characterization of intentional agents, formally defining some more of their properties. Secondly, the mental states that we have characterized don't exhaust the whole category: other mental states need to be specified, like fears, hopes, remembrances and others. Moreover, we have only treated the first-order (in Dennett's terms, see Introduction) behaviour of agents and a second-order analysis is surely needed.

Intentional agents, mental states and mental objects are not all the mental entities we commonly talk about; there are also mental processes and mental events, and these need to be characterized as well.

Finally, we need to relate what is inside an agent with what is outside; this means two things: on the one hand we have to describe the relation holding between mental objects and the "external" objects they represent or stand for, on the other hand we should connect the mental states of an agent with the mental states of other agents, in order to have agents that are able to reason on the mental states of others and to infer from their behaviour what is on their mind, and thus to make predictions on their future behaviour. The analysis of this last topic is particularly important, as it should enable us to connect the mental dimension of ontology with the social one.

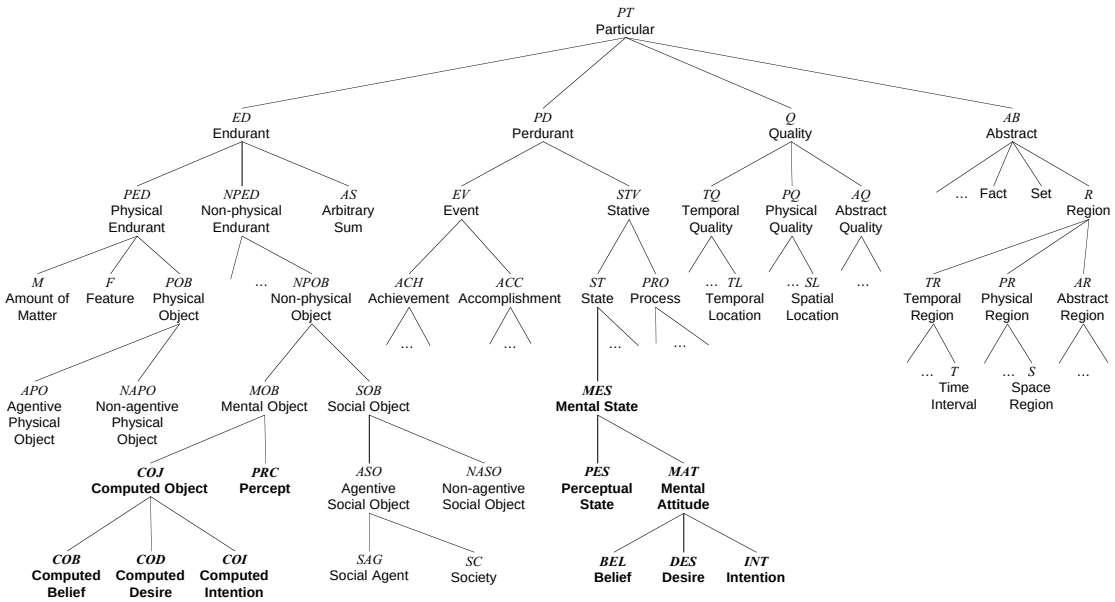

Figure 1

Taxonomy integrating DOLCE and COM basic categories

Acknowledgments

We would like to thank Claudio Masolo and Carola Catenacci for the concrete help and fruitful discussions and comments on the final version of this paper. Without this help a substantial part of the paper wouldn't have been possible. This work was supported by the National project TICCA (Cognitive Technologies for Communication and Cooperation with Artificial Agents).

References

[1] A. Clarke. *Natural Born Cyborgs. Minds, Technologies & The Future of Human Intelligence*, Oxford University Press, New York, 2003.

[2] M. Luck, McBurney P., Preist, C. *Agent Technology: Enabling Next Generation Computing*, AgentLink2003.

[3] C. Masolo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider, "WonderWeb Deliverable D17: The WonderWeb Library of Foundational Ontologies," 2002.

[4] J. A. Fodor. *The Language of Thought*, Thomas Crowell, New York, 1975.

[5] D. C. Dennett. *Brainstorms*, Penguin, London, 1978.

[6] M. Wooldridge. *Reasoning about Rational Agents*, The MIT Press, Cambridge MA, 2000.

[7] D. C. Dennett. *The Intentional Stance*, The MIT Press, Cambridge MA, 1987.

[8] J. L. Pollock. The Phylogeny of Rationality. *Cognitive Science*: 563-588, 1993.

[9] A. S. Rao and M.P. Georgeff, Modeling Rational Agents within a BDI-Architecture. *KR*, 473-484. 1991.

[10] M. E. Bratman. *Intention, plan and practical reason*, Harvard University Press1990.

[11] M. Wooldridge and N. Jennings. Intelligent Agents: Theory and Practice. *The Knowledge Engineering Review*, 2: 115-152, 1995.

[12] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 1: 14-23, 1986.

[13] R. A. Brooks. Intelligence Without Representation. *Artificial Intelligence*: 139-160, 1991a.

[14] R. A. Brooks, Intelligence Without Reasoning. In *Proceedings of IJCAI-91, 12th Int. Joint Conf. on Artificial Intelligence*, 569-595. Sydney, Australia, 1991b.

[15] P. R. Cohen and H. J. Levesque. Intention Is Choice with Commitment. *Artificial Intelligence*: 213-261, 1990.

[16] C. Castelfranchi. Modelling social action for AI agents. *Artificial Intelligence*: 157-182, 1998.

[17] J. Kim, "Events as Property Exemplifications," in M. Brand and D. Walton, Eds, *Action Theory*, pages 159-177. Reidel, Dordrecht, 1976.

[18] H. Steward. *The Ontology of Mind*, Oxford University Press, Oxford, 1997.

[19] T. Crane. *Elements of Mind*, Oxford University press2001.

[20] P. M. Churchland. *Matter and Consciousness*, The MIT Press1988.

[21] G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic*, Routledge, London, 1996.

[22] J. F. Allen and P. J. Hayes, A Common-Sense Theory of Time. In *Proceedings of IJCAI-1985*, 5th *International Joint Conference on Artificial Intelligence (IJCAI85)*, 528-538. Los Angeles, 1985.