EDA: A New Agent Model

Joaquim Filipe
Escola Superior de Tecnologia de Setúbal
Instituto Politécnico de Setúbal
Portugal

and

Paulo Brito
Instituto para os Sistemas e Tecnologias de Informação, Controlo e Comunicação
Lisboa
Portugal

ABSTRACT

Distributed Artificial Intelligence is becoming a mature field. Several agent models have been proposed without a standard being accepted. This paper presents an agent model developed from sound philosophical principles and discusses the coordination and the representation of social structures based on using this agent model.
The EDA model (Epistemic-Deontic-Axiological) is an agent model based on the social psychology theoretical classification of norms and corresponding attitudes: Ontological, Epistemic, Deontic and Axiological. EDA agents are situated in normative information fields, and are described in terms of the basic attitudes aforementioned. Information fields are used as the basis for coordinating an organization, which is seen here as a collective agent composed of other individual and/or collective agents and/or roles, that encompasses multiple embedded information fields. Aspects related to normative knowledge are represented using a combination of deontic logic with agency logic.

Keywords: Agents, Multi-Agent Systems, Formal Models, Social Systems, Logics

1. INTRODUCTION

Organisations are multi-agent systems, eventually including both human and artificial agents. Organisations are also seen as multilayered Information Systems (IS) themselves, including an informal subsystem, a formal subsystem and a technical system as shown in figure 1.

We aim at improving the technical subsystem within the constraints defined by the other two. Organisational information systems are inherently distributed, nowadays, thus communication and coordination are major problems in this kind of information systems. Perhaps motivated by the difficult problems there is currently a strong interest on this area, which is an active research field for several disciplines, including Distributed Artificial Intelligence (DAI), Organisational Semiotics and the Language-Action Perspective, among others. Our approach integrates elements from these three perspectives.
The Epistemic-Deontic-Axiologic (EDA) designation refers to the three main components of the agent structure described in this paper. Here we propose an agent model which, contrary to most DAI proposals, not only accounts for intentionality but is also prepared for social interaction in a multi-agent setting. Existing agent models emphasise an intentional notion of agency – the supposition that agents should be understood primarily in terms of mental concepts such as beliefs, desires and intentions. The BDI model (Rao and Georgeff, 1991) is a paradigm of this kind of agent.
We claim that the cognitive notions that constitute the basis of intentional models show only one face of the coin, the other one being social agency. It is required that an adequate agent model, able to function in a multi-agent setting, should emphasise social agency and co-ordination, within a semiotics framework. Our approach focuses on organisational agents who participate in organisational processes involving the creation and exchange of signs, *i.e.* knowledge sharing, in a way that is inherently public, thus depending on the agent's social context, *i.e.* its information field (Stamper, 1996), to support co-ordinated behaviour. An information field is composed by all the agents and objects involved in a certain activity and sharing a common ontology.

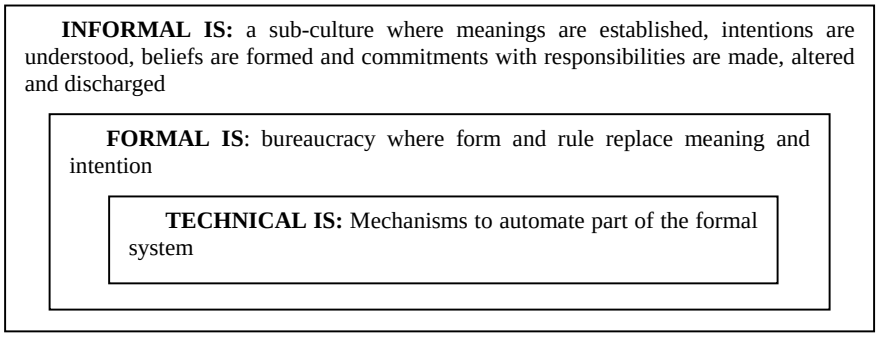

Figure 1: Three main layers of the real information system (Stamper 1996)

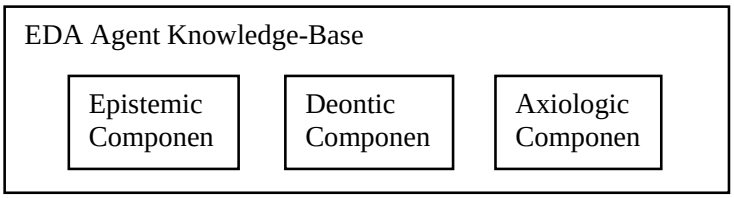

Figure 2: The EDA Agent Knowledge-Base Structure

A realistic social model must be normative: both human agents and correctly designed artificial agents ought to comply with the different kinds of norms that define the information field where they operate, although exceptions may occur. Private representations of this shared, normative, knowledge are translations of the public knowledge into specific agent mental structures. When designing an artificial agent, the designer must adopt a private knowledge representation paradigm to set up the agent's internal knowledge in such a way that it fits the normative shared ontology.

We postulate that norms are the basic building blocks upon which it is possible to build co-ordination among organised entities and co-ordinated actions are the crux of organised behaviour. We claim that although organised activity is possible either with or without communication, it is not possible without a shared set of norms. Therefore, these socially shared norms define an information field that is a necessary condition for heterogeneous multi-agent co-ordination including both artificial agents and humans.

2. THE NORMATIVE STRUCTURE OF THE EDA MODEL

Norms are typically social phenomena. This is not only because they stem from the existence of some community but also because norms are multi-agent objects (Conte and Castelfranchi, 1995):

They concern more than one individual (the information field involves a community)

They express someone's desires and assign tasks to someone else

Norms may be regarded from different points of view, deriving from their social circulation: in each case a norm plays a different cognitive role, be it a simple belief, a goal, a value, or something else.

Social psychology provides a well-known classification of norms, partitioning them into perceptual, evaluative, cognitive and behavioural norms. These four types of norms are associated with four distinct attitudes, respectively (Stamper, 1996):

- Ontological – to acknowledge the existence of something;
- Axiologic – to be disposed in favour or against something in value terms;
- Epistemic – to adopt a degree of belief or disbelief;
- Deontic – to be disposed to act in some way.

An EDA agent is a knowledge-based system whose knowledge base structure is based on the following three components: the Epistemic, the Deontic and the Axiologic.

The epistemic model component is where the knowledge of the agent is stored, in the form of statements that are accepted by that agent. Two types of knowledge are stored here: declarative knowledge – statements relative to the agent beliefs – and procedural knowledge – statements concerning the know-how of the agent, *e.g.* their plans and procedural abilities.

The importance of norms to action has determined the name we have chosen for the model component where the agent goals are represented. An agent goal may be simply understood as the *desire to perform an action* (which would motivate the designation of *conative*) but it can also be understood, especially in a social context, as the *result of the internalisation of a duty or social obligation* (which would motivate the designation of *deontic*). We have adopted the latter designation not only because we want to emphasise the importance of social obligations but also because personal desires can be seen as a form of 'generalised' obligation established by an agent for himself. This provides a *unification* of social and individual drives for action, which simplifies many aspects of the model.

The axiologic model component contains the value system of the agent, namely a partial order that defines the agent preferences with respect to norms. The importance of the agent value system is apparent in situations of conflict, when it is necessary to violate a norm. This preference ordering is dynamic, in the sense that it may change whenever the other internal components of the agent model change, reflecting different beliefs or different goals.

3. THE EDA MODEL INTERNAL ARCHITECTURE

Using the social psychology taxonomy of norms, and based on the assumption that organisational agents' behaviour is determined by the evaluation of deontic norms, given the agent epistemic state, with axiological norms for solving eventual interest conflicts, we propose an intentional agent model, which is decomposed into three main components: the epistemic, the deontic and the axiologic. Additionally there are two external interfaces: an input (perceptual) interface, through which the agent receives and pragmatically interprets messages from the environment and an output (acting) external interface through which the agent acts upon the environment, namely sending messages to other agents[1].

[1] In this paper we restrict our attention to the semiotic, symbolic, types of agent activity, ignoring substantive physical activities.

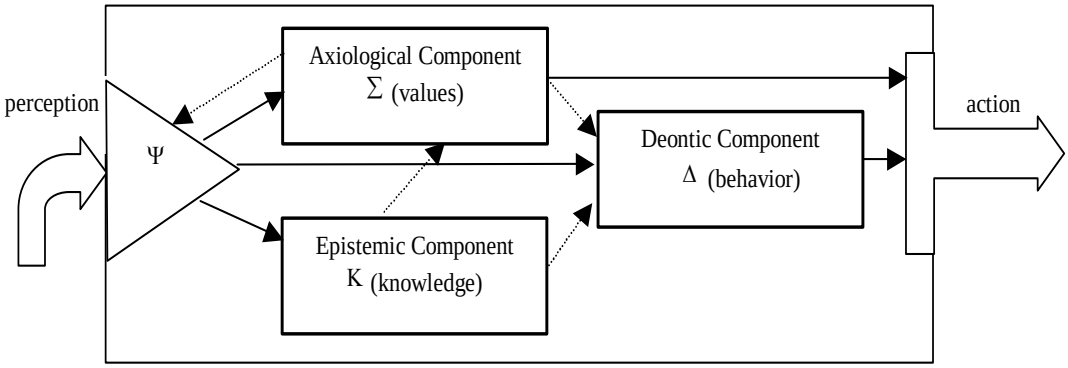

Figure 3: The EDA model component relationships.

A socially shared ontology is partially incorporated in an agent cognitive model whenever it is needed, *i.e.* when the agent needs to perform a particular role. In this case, beliefs are incorporated in the Epistemic component, obligations and responsibilities are incorporated in the Deontic component and values (using a partial order relation of importance) are incorporated in the Axiologic component – all indexed to the particular role that the agent is to play.

Figure 3 depicts the EDA model and its component relationships.

Ψ is a pragmatic function that filters perceptions, according to the agent ontology, using perceptual and axiologic norms, and updates one or more model components.

Σ is an axiologic function that is used mainly in two circumstances: to help decide which signs to perceive, and to help decide which goals to put in the agenda and execute.

K is a knowledge-based component, where the agent stores its beliefs both explicitly and implicitly, in the form of potential deductions based on logical reasoning.

Δ is a set of plans, either explicit or implicit, the agent is interested in and may choose to execute.

The detailed description of each component, including its internal structure, is provided in (Filipe, 2000). In this paper we focus on the system behaviour. The next sections show how in EDA we specify ideal patterns of behaviour and also how we represent and deal with non-ideal behaviours.

4. ORGANISATIONAL MODELLING WITH MULTI-AGENT SYSTEMS USING THE EDA MODEL

The EDA model may apply to both human and artificial agents, and is concerned with the social nature of organisational agents:

Firstly, because it accounts for a particular mental structure (Epistemic-Deontic-Axiologic) that is better, for our purposes, than other agent mental structures proposed in the literature to model agent interaction. Specifically, we intend to use it for modelling information fields where social norms influence individual agents and are used by an agent to guide inter-subjective communication and achieve multi-agent co-ordination.

Secondly, because the model is based on normative notions that are not only intended to guide the agent behaviour in a socially accepted way, but also to identify what sanctions to expect from norm violations, in order to let the agent take decisions about its goals and actions, especially when co-ordination is involved. The EDA model is based on the claim that multi-agent notions such as social commitment, joint intentions, teamwork, negotiation and social roles, would be merely metaphorical if their normative character were not accounted for.

Given its social-enabled nature, the EDA agent notion may be used to model and implement social activities, involving multi-agent co-ordination. However, although the agent paradigm described in this thesis is suited to model team work and joint problem solving, the major novelty with respect to other current agent models is the normative flavour. EDA agents are able to co-ordinate on the basis of shared norms and social commitments. Shared norms are used both for planning and for reasoning about what is expected from other agents but, internally, EDA agents keep an intentional representation of their own goals, beliefs and values.

Co-ordination is based on commitments to execute requested services. Commitments are represented not as joint intentions based on mutual beliefs, as is the case of the Cohen-Levesque model, upon which the BDI paradigm is based, but as first-class social concepts, at an inter-subjective level. The organisational memory embedded in the representation of socially accepted best-practices or patterns of behaviour and the handling of sub-ideal situations is, in our opinion, one of the main contributions that a multi-agent system can bring about.

5. GENERALISED GOALS

Goal-governed behaviour can be represented, according to agency logic (Belnap, 1991), by an agentive statement where an agent α sees to it that the state of affairs *Q* obtains at time τ_2 by the same agent α performing a choice at time τ_1: $[\alpha\ stit : Q]$. We claim that it is possible to use the same syntax and semantics both for representing goals derived from both agent interests and behavioural social norms. This would make it possible to build up detailed plans from agent interests and behavioural norms using means-ends analysis – the same inference mechanism commonly used in knowledge-based systems for goal oriented backward-chaining.

We propose an unification of different concepts under the notion of generalised goal. This unification stems from the following axioms:

All actions have a responsible agent, who performs the action.
The performer may not be the ultimate responsible for the action and, in the case of artificial agents, it never is: the last agent in the responsibility chain must be a person.
An action is performed for satisfying someone's interest.
If the client for whom the action is performed is not satisfied he can ask for a sanction to be imposed by the information field that defines the context of the action.
We propose the following formal representation for generalised goals:

$G_i = O^{\beta/\chi}_{\alpha/\rho}(Q,\tau,\sigma) \supset O$ ($[\alpha\ stit : Q]$ in-time-window τ subject-to-sanction σ)

This can be read as: α (under the responsibility of ρ) has the goal of ensuring Q, in reply to a request of β (under the control of χ); α is the performing agent, β is the client that requested the service, χ is a controller agent – which Castelfranchi (1993) calls a witness, Singh (1996) calls a context, and we associate to the information field. ρ is the responsibility chain for α [2]; O is the standard deontic operator 'ought-to-be'; $[\alpha\ stit : Q]$ is an agency statement, saying that agent α sees-to-it-that proposition Q becomes true. This means that α will perform a plan to bring about Q in time window τ, where τ is a time expression specifying the time window during which proposition Q is intended to be satisfied; τ may be specified in absolute time or relative to some event; σ indicates the sanction cost of violation.
If agent α is different from agent β then we have a social commitment; otherwise we have an intention. In the latter case there is no controller agent χ and the sanction σ is typically null (represented by $\varnothing$) with expected value zero.

6. CONCLUSIONS

The EDA model is a norm-based, theroretically sound, agent model that takes into account not only the intentional aspects of agency but also the social norms that prescribe and proscribe certain agent patterns of behaviour. The main components of this model (Epistemic, Deontic and Axiological) have a direct relationship with the types of norms that are proposed in the social psychology theory supporting the model. In this paper, however, we focused our attention essentially in the Deontic component, where the normative social aspects are more important, namely where ideal and sub-ideal behaviours are represented.
We consider agents to be goal-governed systems. Using the *generalised goal* concept, all agent goals can be represented as obligations, encompassing both agent self-imposed obligations and social obligations – derived from moral obligations or commitments established in the course of their social activity. The proposed modal operator – G_i – overcomes the excessive strength of *stit*, and implicitly permits the representation of an important organisational activity: delegation, which Belnap's *stit* does not allow. Furthermore, the inclusion of a '*committed-to'* agent also extends *stit* and permits the bringing in of the notion of social commitment, based on the notion of obligation towards other agents and subject to sanctions.
Instead of manipulating the notions of belief, intention, desire, commitment, goal and obligation, the only primitives that we need are *belief* and *generalised goal*. All other notions are constructed from these ones. The interested reader can find more details about this in (Filipe, 2000).

[2] The responsibility chain is a list constituted by one or more agents that are responsible for the commitment to service performance and may be brought to scene if there is a breakdown during the service performance or if the performing agent is unable to handle the service satisfactorily.

7. REFERENCES

Belnap, N., 1991. Backwards and Forwards in the Modal Logic of Agency. Philosophy and Phenomenological Research, vol. 51.

Carmo, J. and A. Jones, 1998. Deontic Logic and Contrary to Duties. Personal copy of a pre-print manuscript.

Castelfranchi, C., 1993. Commitments: from Individual Intentions to Groups and Organisations. In Working Notes of AAAI'93 *Workshop on AI and Theory of Groups and Organisations: Conceptual and Empirical Research*, pp.35-41.

Cohen, P. and H. Levesque, 1990. Intention is Choice with Commitment. *Artificial Intelligence*, 42:213-261.

Conte, R., and C. Castelfranchi, 1995. *Cognitive and Social Action*, UCL Press, London.

Dignum, F., J-J Ch. Meyer and R.Wieringa, 1994. A Dynamic Logic for Reasoning About Sub-Ideal States. In J.Breuker (Ed.), *ECAI Workshop on Artificial Normative Reasoning*, pp.79-92, Amsterdam, The Netherlands.

Filipe, J., 2000. Normative Organisational Modelling using Intelligent Multi-Agent Systems. Ph.D. thesis, University of Staffordshire, UK.

Hilpinen, R., 1971. *Deontic Logic: Introductory and Systematic Readings*. Reidel Publishing Co., Dordrecht, The Netherlands.

Rao, A. and M. Georgeff, 1991. Modeling Rational Agents within a BDI architecture. *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning* (Nebel, Rich, and Swartout, Eds.) pp.439-449. Morgan Kaufman, San Mateo, USA.

Stamper, R., 1996. Signs, Information, Norms and Systems. In Holmqvist et al. (Eds.), Signs of Work, Semiosis and Information Processing in Organizations, Walter de Gruyter, Berlin, New York.

Singh, M., 1990. *Towards a Theory of Situated Know-How*. 9th European Conference on Artificial Intelligence.

Singh, M., 1996. Multiagent Systems as Spheres of Commitment. Proceedings of the International Conference on Multiagent Systems (ICMAS) - Workshop on Norms, Obligations and Conventions. Kyoto, Japan.

Werner, E., 1989. Cooperating Agents: A Unified Theory of Communication and Social Structure. In Gasser and Huhns (Eds.), *Distributed Artificial Intelligence*, pp.3-36, Morgan Kaufman, San Mateo, USA.

von Wright, G., 1951. Deontic Logic. Mind, 60, pp.1-15.