On the Metaphysics of Agents

Paul Davidsson
pdv@bth.se

Stefan J. Johansson
sja@bth.se

Department of Systems and Software Engineering
School of Engineering
Blekinge Institute of Technology
372 25 Ronneby, Sweden

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence

General Terms

Theory, Measurement, Documentation

Keywords

Agent definitions, Agent properties

1. INTRODUCTION

Agent technology has not been accepted by software developers to the extent and rate anticipated. We believe that the lack of clarity and consistency regarding the terminology and its use may be a contributing cause for this. The dissemination of research results both to industry and related research disciplines becomes difficult. In addition, a typical sign of a mature scientific field is that the interpretation of the central concepts has converged. Although, this process is complicated by the fact that agent researchers constitute a heterogeneous group with different backgrounds, we believe that a common understanding of core concepts is a prerequisite for a broad industrial acceptance. In order to achieve this, we need to study the very nature of agents, or, in other words, the *metaphysics* of agents. We will in this work: (i) analyze the definitions of *agents* currently used, (ii) investigate whether these definitions correspond to how the term is actually used by researchers, and (iii) propose an improved definition of agents.

2. AGENT AND PROPERTY DEFINITIONS

We have chosen to focus on some of the most influential definitions of agents. They are taken from books that either are widely used in education [2, 5, 7, 8], or is a recognized state-of-the-art survey [4]. In order to compare the definitions, a set of agent properties will be defined. The property definitions are based on existing definitions. We describe how we interpret the definitions when judging whether an agent in a particular study has the property or not.

Perceptive is the ability to classify and distinguish states of the world, not only with respect to prominent features of the environment, but also with respect to the actions it is undertaking [2].

AAMAS'05, July 25-29, 2005, Utrecht, Netherlands.

Acting is the property of being able to transform the state of the system by modifying the positions of and the relationships existing between the objects (of the system) [2].

Situated is the capability for an agent in an environment to recognizing the objects situated in the environment through the function of its perceptive capabilities and of transforming the state of the system by modifying the positions of and the relationships existing between the objects [2]. We will from here on consider situated as being composed of the two properties perceptive and acting.

Autonomy is a more complex concept, often defined as the ability of exercising choice over their actions and interactions [4]. Verhagen [6] has identified four levels of autonomy: Reactive, Plan autonomous, Goal autonomous, and Norm autonomous, which we will use in our analysis. The following assumptions are made: All agents that are acting are at least *reactive*. *Plan autonomous agents* cannot change their goals, but are able to choose actions from a repertoire and perform them in an order that will take them closer to their goals. For an agent to be *goal autonomous*, we require it to be able to consider, reason about, and change its goals (cf. proactivity). Finally, a *norm autonomous agent* is able to choose which goals are reasonable to pursue, based on its system of norms. These norms can be changed, e.g. as a consequence of a goal conflict.

An agent that is *(weakly) rational* tries to satisfy its objectives by taking account of the resources and skills available to it and depending on its perception, its representations and the communication it receives [2].

Communicative describes the capability of performing communicative actions in an attempt to influence other agents appropriately [8].

Some properties used in the definitions may be very hard to decide objectively. We argue that *Problem solving* [4] and *Service providing* [2] belong to this category. Whether an agent solves a problem or provides a service depends on the environment in which it acts and who is observing it. It can therefore be argued that these properties are unsuitable in defining an agent as they describe the system level rather than the agent. Moreover, as the interesting aspects of Problem solving and Service providing are captured by other properties like Goal driven, Rational, and in the second case, Communicative, we have chosen to disregard them in the analysis. We have also chosen to exclude *computerized, computational,* etc. since all investigated papers exhibit such properties (and all definitions, except [5], assume them). Since agents that are at least plan autonomous are goal driven, we also exclude this property in the further analysis.

3. ANALYSIS

We have analyzed the definitions in terms of the properties de-

Property	[2]	[8]	[7]	[4]	[5]	% of Agents
Perceptive	E	I	E	I	E	96
Acting	E	I	E	I	E	94
Weakly rational	I			I		97
Communicative	E					81
Reactive	I	E	E	E		29
Plan Autonomous	I	E	E	E		37
Goal Autonomous						25
Norm Autonomous						9

Table 1: The properties of the different definitions of agents and the percentage of papers having agents with them. E means that the property is necessary and explicit in the definition, I that it is necessary and implicit.

fined in the last section. Table 1 summarize the relations between the agent definitions and agent properties.

We see that the definition of Russell and Norvig is the most general, requiring only two properties (perceptive and acting). The definition of Ferber can be seen as the most specific as it requires the largest number of properties. Note that it does not explicitly require autonomy, which is required by all others, except for Russell and Norvig.

Using the agent properties described above, we have tried to determine how each full paper in the AAMAS 2003 proceedings use the term agent. Table 1 also shows how many of the papers that assume their agents to have the different properties (or, for application papers, actually have them). The detailed results of the analysis can be found in [1].

The definition of Ferber does not mirror the actual usage very well (18 out of 115 papers). On the other hand, the definition of Russell and Norvig is consistent with almost all of the papers (108 papers), whereas the definitions of Wooldridge and Weiss are consistent with 79 of the papers.

The definition by Russell and Norvig is consistent with almost all of the papers. But, is this necessarily a good thing? Perhaps the definition is too general to be meaningful, covering also entities that there is little value in regarding as agent, possibly even contributing to undesired confusion. As pointed out by Franklin and Graesser [3]: "If we define the environment as whatever provides input and receives output, and take receiving input to be sensing and producing output to be acting, every program is an agent."

4. CONCLUSIONS

The main conclusions are: that none of the necessary (but not sufficient) properties used in the definitions were met by all papers, and that there is some correlation between the properties assumed and in which session the paper was presented, e.g., papers in the Agent-Oriented Software Engineering session often assumed a high level of agent autonomy

One way of summarizing the finding of our study is to provide a descriptive (in contrast to prescriptive) definition of the term agent based on how it is used in the papers reviewed. There are three properties that are common for almost all papers, namely *Situatedness*, *Rationality,* and *(plan) Autonomy*. Thus, an agent would in this case be something that is situated, rational and autonomous to some extent. However, we think that it is more useful if we restrict the definition to software agents (seeing robotic agents as embedded software agents) as all the papers deal with this type of agents.

Since software agents often are argued to be en extension of the concept of software objects, as used in object-oriented programming, we may build upon the definition of the term software object. A typical definition is "an object is a self-contained entity that consists of both data and procedures to manipulate the data" (http://www.pcwebopaedia.com/). It can be argued that a software agent shares the property of being self-contained and containing data. However, it has no procedures for manipulating that data, rather it has the ability to perceive and act in an environment, which we here refer to as being situated. Also, it seems more reasonable to talk about states than data when it comes to agents. Thus, an agent could be defined as: *a self-contained entity that has a state, which is situated (able to perceive and act) in an environment, rational, and at least reactively autonomous.* This is a definition that covers nearly all of the papers, but also excludes many entities that typically are not regarded as agents.

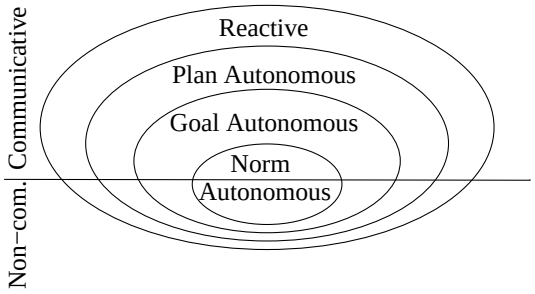

Figure 1: Descriptive classification of agent types.

This definition would then correspond to a basic "vanilla" agent. In addition, it could be useful to talk about different types of agents. Based on our analysis, we identify two properties that could be very useful for defining different types of agent, namely Autonomy and Communicative. As illustrated in Figure 1, we may classify agents according to their degree of autonomy (from reactive to norm autonomous) and whether they are communicative or not.

Another possibility that we will investigate is that on having different definitions of agents on different levels of abstraction. For instance, one definition of what an agent is on the conceptual/modeling level, and another one explaining what it means to be an agent on the level of implementation.

5. REFERENCES

[1] P. Davidsson and S. Johansson. On the metaphysics of agents (extended version). Research Report 2005:04, Blekinge Institute of Technology, 2005.

[2] J. Ferber. *Multi-Agent Systems, - An Introduction to Distributed Artificial Intelligence*. Addison W., 1995.

[3] S. Franklin and A. Graesser. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In *Intelligent Agents III. Agent Theories, Architectures and Languages (ATAL'96)*, volume 1193 of *LNAI*, Berlin, Germany, 1996. Springer-Verlag.

[4] M. Luck, P. McBurnley, and C. Preist. *Agent Technology: Enabling Next Generation Computing - A Roadmap for Agent Based Computing*. AgentLink, 2003. ISBN 0854 327886.

[5] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2003. 2nd edition.

[6] H. Verhagen. *Norm Autonomous Agents*. PhD thesis, Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden, 2000.

[7] G. Weiß. *Multiagent Systems, – a modern approach to distributed artificial intelligence*. MIT Press, 1999.

[8] M. Wooldridge. *An introduction to Multi-Agent Systems*. Wiley, 2002.