2001

Describing intelligent agent behaviors

Walt Truszkowski
Goddard Space Flight Center

Jagannathan V. Iyengar
University of Wisconsin

Fred Mills
Bowie State University

Describing intelligent agent behaviors

Walt Truszkowski
Goddard Space Flight Center

Jagannathan V. Iyengar
University of Wisconsin-Whitewater

Fred Mills
Bowie State University

ABSTRACT

The development of new intelligent agents requires an interdisciplinary approach to programming. The initial challenge is to describe the desired agent behaviors and abilities without necessarily committing the agent development project to one particular programming language. What are the appropriate linguistic and logical tools for creating a top level, unambiguous, program-independent, and consistent description of the functions and behaviors of the agent? And how can that description then be translated easily into one of a number of program languages? This article provides a case study of the application of a simple Belief, Desire, and Intention (BDI) first order logic to a complex set of agent functions of a theoretical community of intelligent nano-spacecraft. The basic research was conducted at NASA-GSFC (Greenbelt), Advanced Architecture Branch, during the summer of 2001. The simple examples of applied BDI logic presented here suggest broad application in agent software development.

INTRODUCTION

The demand for intelligent agent software is likely to grow as both public and private sector innovators seek to deploy adaptive, autonomous, information technologies to production, scheduling, resource management, office assistance, information collection, remote sensing, and other complex functions. Intelligent agents, also knows as autonomous agents, are distinguished from other software programs by their ability to respond to a changing environment in pursuit of goals.

One of the most critical stages of such intelligent agent development is the basic research that goes into determining how to translate client needs into agent software. Since programmers and operations managers have different expertise, there is often a linguistic gap between the functional language of the customer and the technical programming language of the agent developer. If the customer and agent developer do not speak a common language, both time and money may be lost in needless errors and misunderstandings.

99

There is a need for a program-independent, unambiguous, logically consistent language for describing the desired computer agent practical reasoning abilities and behaviors. Program independence gives the customer maximum flexibility in choosing a programming language to implement the desired functions. The avoidance of ambiguity allows the customer to say exactly what she means. And logical consistency ensures that only valid arguments will be generated by the agent's knowledge base, avoiding false beliefs about the world from being deduced from true beliefs.

A program independent language for describing agent functions, practical reasoning, and behaviors, solves what may be called the *top-level description problem*. The top level of description should be close enough to ordinary English to be interdisciplinary, yet rigorous enough to avoid ambiguity and inconsistency. The strategy of solving this problem borrows from first order logic, cognitive science and recent work in the field of artificial intelligence, specifically autonomous agent theory (see, e.g., the collection of articles in Muller, Wooldridge, & Jennings, 1997).

The solution to the top level description problem suggested by some of the recent autonomous agent research is to combine the practical reasoning tools developed by psychologist Michael Bratman's belief, desire, and intention (BDI) framework (1987), with some of the tools of first order logic. This hybrid language is usually referred to as BDI logic. There are a number of BDI logics under development; here we employ the Logic of Rational Agents (LORA) developed by Michael Wooldridge (2000) to illustrate how BDI logic works in practice.

The authors faced the top-level description problem at NASA during the basic research stage of specifying the behaviors of a community of agents, in this case, the Autonomous Nano-Technology Swarm (ANTS).[i] ANTS will be designed to engage in practical reasoning and behaviors that implement the variety of functions necessary to explore the asteroid belt and communicate observations to earth.

---

[i] The ANTS concept is indebted to the work of Mark Campbell, et al. (05/31/99) in "Intelligent Satellite Teams for Space Systems." As an example of an Intelligent Satellite Team (IST), the report considered "a large number of satellites orbiting the Sun using optical and infrared sensors to look for asteroids" (p. 7). These satellites would share data, track, and collect information about interesting asteroids. The report mentions the important role of practical reasoning in IST: "Reasoning in an agent allows the agent to reach an intelligent conclusion or decision about a given set of data" (p. 31). The report notes that this practical reasoning would give the agent more adaptability and autonomy. But there is no attempt to develop a top-level language for describing this type of reasoning.

In another foundational paper, Steven Curtis, et al. (2000) provides a description of the mission, system architectures, and ANT functions, including the type of instruments ANTS will employ to collect data bout asteroids. There are also descriptions of how ANTS might coordinate their functions. The operational scenarios, however, do not attempt to represent the state of knowledge of each ANT, and therefore do not reveal the anatomy of practical reasoning. Again, a language of ANT practical reasoning would be required to carry out such an objective.

Turszkowski, Zoch, and Smith (2000) point out that a subsystem agent must be able to exercise some "reasoning process" in order to make use of external resources in new ways (p. 6). The first stage in realizing spacecraft constellation autonomy is "agent development." We suggest that this stage requires an economic way to describe practical reasoning of the agents.

METHODOLOGY

The first methodological consideration is to determine the level of detail that is appropriate in applying LORA as an agent development tool. The purpose of using BDI logic here is not to prove theorems; a sufficient body of deductive proofs has already been developed in first order logic textbooks. Nor is it to impress the client with obtuse descriptions of functions. It is rather to use just enough symbolism to economize the clarify the practical reasoning of the agents.

The second methodological consideration is to announce our position on exactly what we mean to imply when employing psychological terms to describe computer agents. In what John Searle calls strong AI, conscious experience is attributed to intelligent computer agents (1984). In this paper we remain agnostic with regard to the debate over whether computer agents have real as opposed to "as if" intentionality (see Mills, 1998a; 1998b, for a detailed discussion). We discuss only those features of the debate that elucidate the advantages of using BDI logic as opposed to engineering or design idioms for top-level descriptions of communities of agents.

Part one of this paper provides theoretical justification for the use of Bratman's BDI framework for understanding agent behaviors. Part two provides an explanation of how first order logic can be combined with BDI and a dynamic component to account for agent decisions in time. Part three presents a practical problem of describing the behavior of a community of agents involved in a very complex scientific endeavor, in this case, the ANTS exploration of the asteroid belt. We believe the insights presented here have practical implications for the large variety of agents that will be developed in the near future.

PART I. BRATMAN'S BDI FRAMEWORK: THE INTENTIONAL STANCE

The intentional stance treats a sufficiently complex system as if it had mental states and engaged in practical reasoning. The term "intentionality," first used in empirical psychology by Franz Brentano (1973/1874), means directness towards an object. As John Searle points out (1984), one of the unique features of mental states is that they are always directed towards or about some object. If I perceive, I perceive something. If I believe, I believe that something is the case. If I intend something, I have some purpose in mind. We often use the intentional stance towards artifacts also, as when we say the car does not want to start. But we do not really believe that a car has desires; we speak only of the car "as if" it had desires to avoid a mechanical description of what is going on when it fails to start.

Regardless of where one stands in the philosophical debate regarding whether more complex machines such as computer agents have real or just "as if" intentionality, both sides generally agree that the intentional stance is an economic way to describe a complex system that engages in practical reasoning. Daniel Dennett (1978) points out that this economy of expression becomes clear when we contrast the intentional stance with the engineering stance. ANTS provides a good example of the economy of expression that results from opting for the intentional stance for top-level descriptions of agent behaviors.

101

Imagine a community of nano-spacecraft setting out to explore the asteroid belt. At some point in the deployment of this community, ANT #24 determines some activity is a priority and desires to execute that activity. At the programming level of description one would need to know the semantics and syntax of the programming language and read the code. At the even more detailed engineering level, one would then translate this higher level code into the impossibly complex machine code.

By contrast, if one employs the intentional stance, the practical reasoning of Agent #24 becomes simple and concise: "Agent #24 *desires* to observe Asteroid Psyche, *believes* that the conditions are now optimal for such observation, and thus *intends* to make such observations." The next step is to imbed this intentional language in a first order logic to ensure the consistency of all descriptions of the ANTS behaviors.

PART II. COMBINING DBI WITH FIRST ORDER LOGIC AND DYNAMICS

First Order Logic

First order logic contains simple symbolic rules for combining propositions. For example, "Anna is a student and Anna is a biology major" can be represented by conjunction: $p \wedge q$. First order logic provides a generally accepted consistent logic. It is ideal for stating combinations and relations between beliefs, desires, and intentions about the world. It allows us to state just what the system knows without having to account for what it does not yet know (see Levesque & Lakemeyer, 2000, for a discussion).

In order to represent the knowledge of the agent, some initial set of beliefs about the world are formulated. The inference of rules of first order logic are used to ensure that the beliefs are consistent and that arguments are always valid. By following the basic rules of first order logic, new beliefs can be deduced from the current set of beliefs and acquired beliefs. (These acquired beliefs are based, e.g., on new inputs from the environment.)

In the case of ANTS, the initial beliefs are about known asteroid types, relative locations, shapes, rotations, mass, distribution, gravity, albedo and in some cases provisional classifications of asteroids (P. E. Clark, ANTS 04/05/01). The inputs will originate in sensor instrument data and communications data.

Example in ANTS Knowledge Base

An example of a representation in the ANTS knowledge base using some simplified LORA is:

- (differentiated Psyche ∨ (undiff) Psyche   [either,or]
- (differentiated Psyche ⇒ (PriorityObject) Psyche   [if, then]
- (differentiated)Psyche {new knowledge!}
- (5000MIFRAnt#24) Psyche

102

Notice that first order logic is used here, without BDI, to represent the truism that asteroid Psyche is either differentiated or undifferentiated. If it is differentiated, it becomes a priority object of investigation that could trigger other actions. Assume that sensor information confirms that Psyche is differentiated. It then follows necessarily that Psyche becomes a priority object. The last formula states that ANT #24 is in the proximity of Psyche. We can expect that, given the right combination of BDI, #24 will set as a goal the closer observation of Psyche and finally actuate behaviors towards the attainment of the goal.

In order to begin the construction of a language that includes practical reasoning, Wooldridge adds a Belief component to represent epistemic states (2000). Thus we refine our description of ANT #24's knowledge by adding:

$$(Bel \; q \; PriorityObject(Psyche))$$

Here, agent #24 (represented by the symbol "q") believes that psyche is a priority object. This belief can be updated or modified given additional information from the sensors or new communications.

In order to complete Bratman's framework for practical reasoning, we now add the intentional and emotive components: intentions and desires. Consider a desire as an enduring goal. An agent may have a number of desires, only some of which are realizable at any given time. An intention is a pro-attitude. That is, it is a movement toward an immediate objective until that objective is fulfilled or some new event changes the current intention (Wooldridge, 2000). This distinction between intentions and desires helps us to explain how an agent may change course in response to environmental variables that change through time. Here is an example of epistemic and emotive states combined, using LORA:

- (Des ruler (Bel worker#24 PriorityObject (Psyche)))

Here the ruler agent desires that ANT #24 believes that psyche is a priority object. This mental state is important because it may trigger an action by ANT#24, which has among its desires a desire to pursue priorities set by the ruler ANT.

Dynamic (Temporal) Component

Now that first order logic and BDI have been combined there is one more critical step required to complete a top-level description of autonomous agency: dynamics. In order to represent practical reasoning and resulting behaviors through time, a schema of relevant features of the present state of affairs is constructed. Since not every feature of the universe can be represented, the state of affairs, that is, the world of the agent, contains beliefs about only those features of the universe relevant to the functions of the agent. Since some agent technologies will be able to learn from their experience, in some architecture the relevant states of affairs can expand and change. For the present purpose humans will define the relevant state of affairs. Each node in the schema will branch out into possible worlds. These possible worlds are alternative paths that the agent may select, depending on its current beliefs, desires, and intentions. The following map illustrates a world time schema:

THE WORLD OF ANT

**Each world/time state has a number of
possible futures. a = transition actions**
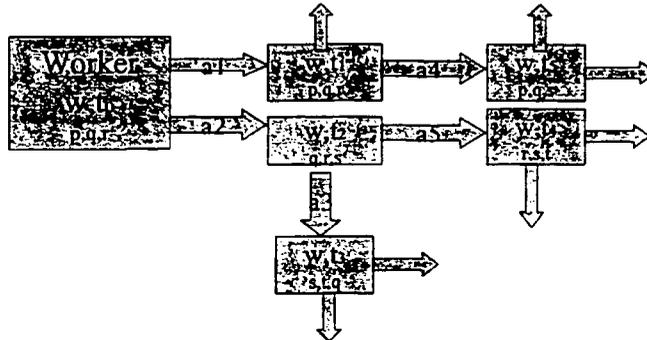
w,t = world,time; p,q,r,s,t are beliefs; a = transition action; branches represent possible worlds (see Wooldridge, 2000 for a more detailed description of agent behaviors through time; the above chart is adapted from pp. 56, 64, 94)

Specification of Functions

We are not ready to specify the functions of various agents in the community of agents and then apply LORA to describe their behaviors.

- General ANT community functions
  - mapping and close-up imaging of Asteroids using multi-spectral band coverage [Clark, Memo 04/05/01]

- Worker ANT functions
  - Communications, resource management, navigation, local status (housekeeping), local conflict resolutions, science data acquisition processing [Curtis et al., 2000, 3]

- Ruler ANT functions
  - Plan assignments for worker ANTS
  - Maintain shared SWARM statistics
  - Resource management [see S. Curtis, et al., 2000, 3]

The ANTS is conceptualized here as a community of autonomous rational agents whose behaviors are generated by a knowledge base (KB), a set of goals, inference procedures, and percepts. Although each worker ANT is autonomous in terms of its own function, it is subordinate to the function of the ruler. The intentions of the ruler are passed on to the workers through the messengers and the workers actuate the plans that achieve the goal.

Each worker has as its permanent goal the appropriate and timely collection or discovery of data from the target type of objects, the maintenance of health and safety, and the timely communication of data to the messengers, who in turn report to the ruler and to earth. The workers' goals are sub-goals of the ruler's goal and the worker's plan at any given state of affairs is a sub-plan of the ruler's plan.

Example of a LORA Specification

In the scenario, the Ruler has received information about an opportunity to view Psyche under ideal conditions and there is a group of worker ants in the neighborhood of Asteroid Psyche (A).

The Ruler forms an *intention* to attain a goal as a result of a deduction that employs beliefs in its KB, acquired beliefs derived from current percepts, and its belief derived from communications with all ANTS. There is a potential for a group of workers and allied messengers to achieve the goal. This is exactly the sort of scenario supported by BDI type logic!

For simplicity temporal considerations, quantification, proofs, and goal/sub-goal relations are left out. We seek to illustrate here the usefulness of BDI logic to model the practical reasoning or "mental states" of cooperating agents.

A = constant for Psyche
PfC = perception of the potential for cooperation
i = ruler agent
j = messenger agent
g = group of worker agents, each with specialized observation instruments
α = action
φ = goal to study Asteroid Psyche
achvs = attain through an action
Bel = Belief
Des = Desire
π = Set of plans to be actuated (executed) by each worker to attain φ
ψ = Preconditions for φ to become the next goal (Science agenda from humans, combination of percepts in relation to KB)
J-attempt = joint attempt

Assumption: The Ruler has formed an intention to achieve the goal of collecting data about Asteroid Psyche. We do not here represent the deliberations that lead to this intention. We begin with the process of mobilizing the workers to achieve the goal and a description of the mental state of the community of agents.

The Ruler forms an intention to attain a goal (to study Psyche) as a result of φ having been met. φ, in this case, is a combination of the ruler's KB, mission priorities, current percepts, its belief that the goal has not yet been attained, its belief that the ruler itself cannot or does not desire to achieve the goal by itself, and its belief that there is a potential for a group of workers and

Journal of International Information Management   Volume 10, Number 2

allied messengers to achieve the goal. The ruler will therefore intend not only the goal, but that the messenger intends the goal and that the group of workers intends the goal and finally achieves the goal.

Additional Specifications

(Int i φ) ^ (PfCi φ) ^ (Int i (Int j φ (Int g φ)))

The ruler intends the goal of studying the Asteroid Psyche and acknowledges the potential for cooperation among the workers to being about the goal and intends that the messenger intends that the group of workers intends the goal.

Int i (Achvs g α π φ)

The ruler intends that each member of the group following its share of the plan can achieve the goal through the group's actions.

{Inform i j α π φ} ^ {Request Thi i α {Inform j g α π φ}}

Agree j i α'}; α'

Here the α is the informing action and α' is the action that j agrees to perform.

{Inform j g α π φ} ^ {Requests Th j g α (Happens α')}

It is assumed that, unless the workers receive percepts, in relation to their KB, that requires that the goal be reevaluated, the group will believe that it can execute the plan to achieve the goal.

(M-Bel g $\neg$ φ) ^ (M-Des(Achvs α π φ)) ^ (M-Int g (Achvs α; φ?)) = {j-Attempt g α φ}

The advantage of using a language like LORA is that we can always add new functions to an agent and make agent behaviors depend on previous knowledge (the set of beliefs in KB and new percepts). We can also add to a list of possible behaviors by adding new beliefs. Most important, we can very economically describe the behavior of the system at a "high" level without employing a program language.

REFERENCES

Braman, M. (1987). *Intentions, plans, and practical reason.* Cambridge: Harvard University Press.

Campbell, M. Principle Investigator. (05/31/99). Intelligent satellite teams for space systems. *Final Report for the NASA Institute for Advanced Concepts Program.* Seattle, Washington.

Clark, P. (04/05/01). Asteroid science for ANTS. Memo to M. Rilee.

Curtis, S., et al. (2000). ANTS (Autonomous Nano Technology Sarem): An artificial intelligent

106

approach to asteroid belt resource exploration. *American institute of Aeronautics and Astronautics.*

Dennett, D. (1978). *Brainstorms.* Cambridge, Massachusetts: The MIT Press/A Bradford Book.

Brentano, F. (1973). *Psychology from an empirical standpoint.* New York: Routledge. (Original work published in 1874.)

Iyengar, J. V. (2001, July). *Intelligent software agents: Human agent interface considerations.* Summer Faculty Fellowship Project. NASA-GSFC.

Levesque, H. & Lakemeyer, G. (2000). *The logic of knowledge bases.* Cambridge, MA: The MIT Press.

Mills, F. (1998a). The easy and hard problems of consciousness: A Cartesian perspective. *The Journal of Mind and Behavior, 2,* 119-140.

Mills, F. (1998b). A critique of the concept of computer moral agency. *Humanities and technology review, 17,* 26-39.

Mills, F. (2001, July). *Rational agents: Using beliefs, desires, and intentions framework for ANTS simulation.* Summer Faculty Fellowship Project. NASA-GSFC.

Muller, J., Wooldridge, M., & Jennings, N. (1997). Intelligent agents III: Agent theories, architectures, and languages. *ECAI '96 Workshop (ATAL) Budapest, Hungary, August 12-13, 1996, Proceedings.* Berlin: Springer-Verlag.

Truszkowski, W., Zoch, D., & Smith, D. (2000). Autonomy for constellations. *SpaceOps 2000 Conference.* France.

Wooldridge, M. (2000). *Reasoning about rational agents.* Cambridge, MA: MIT Press.