From: FLAIRS-01 Proceedings. Copyright © 2001, AAAI (www.aaai.org). All rights reserved.

ATTITUDES FOR AGENTS IN DYNAMIC WORLDS

S. Au N. Parameswaran
School of Computer Science and Engineering,
The University of New South Wales
Sydney 2052 Australia
(sherlock, paramesh)@cse.unsw.edu.au

Abstract

In this paper, we propose that in multiagent dynamic worlds, agents which are autonomous, need to be guided by attitudes for effective problem solving. Often agents, existing in dynamic and hostile worlds, are required to exhibit the specified problem solving behaviors for prolonged periods of time either continually or intermittently. In order to be able to perform these types of behaviors, autonomous agents need meta-level controls known as attitudes to guide them towards selecting the proper goals and actions. In this paper, we investigate the role of attitudes in problem solving in dynamic worlds, and suggest several attitudes for the agents in a hostile dynamic world, the fire world. We then evaluate and compare the problem solving behaviors of the agents in a simulated fire world using different types of attitudes.

In a dynamic environment, meta-level control is needed not only to improve the efficiency of reasoning, but also the accuracy and utility of the results of reasoning. - Martha Pollack in *The uses of plans*, AI 57(1992) 43-68.

Introduction

We consider plan based agents where the agents attempt to solve a problem by deriving a sequence of actions called a plan, and then execute the plan. In a static world (that is, no changes take place when the agent does not perform any actions), a plan based agent, hopefully, is certain to achieve its goal as long as it chooses the right sequence of actions to execute. In dynamic worlds however, even plans that are proved to be correct at the beginning may fail due to unexpected changes in the world. In this paper, we propose that agents which are autonomous, need to be guided by attitudes for effective problem solving in dynamic worlds.

What is an Attitude?

Attitude is "probably the most distinctive and indispensable concept in contemporary American social psychology" [Fishbein 1975, Allport 1935]. However, despite this, there is no clear definitions of what exactly an attitude is and what role it plays in determining the behaviors of humans.

An English Dictionary defines attitude as *a mental view or disposition, esp. as it indicates opinion or allegiance* [Dict 1992]. In social psychology, attitude is defined in several different ways. The meaning that we adopt in this paper, which is also the most popular one, defines attitude as a *learned predisposition to respond in a consistently favorable or unfavorable manner with respect to a given object* [Fishbein 1975]. There are three basic features to attitudes: the notion that it is *learned*, that it *predisposes* action and that such actions are *consistently favorable* or *unfavorable* toward the object. In this paper, we take this as the basis of our definition of attitude, discuss its utility, and finally evaluate the effect of attitudes in problem solving in a hostile dynamic world, viz., the fire world. We discuss attitudes in depth in section 2. In section 3, we define attitudes that are relevant to the fire world. Section 4 presents problem solving with attitudes. We present the performance results of attitude based agents in section 5. In section 6, we review the related work and compare attitudes with other mental attributes and conclude the paper.

Attitudes

In multiagent dynamic worlds, an activity of an agent may often fail. In fact, failures are so common that they may be considered more as rules than as exceptions. An unexpected event can occur for many reasons and cause an immediate failure. However, there are also unexpected situations which may not cause an immediate failure. While this problem is generally difficult for the agent to deal with, we recommend that agents use a set of predefined strategies to deal with these unexpected situations. Certain types of predefined strategies can be classified as attitudes.

We define the attitude of an agent as a *built-in predisposition to respond in a consistently favorable or unfavorable manner with respect to a given object.* In this definition, we adopt a weaker model of attitude than the one define in social physiology and do not insist that an attitude be learned from its past experience. All attitudes an agent possesses are built into the agent by the designer explicitly. An agent may perform different behaviors with respect to an object at different points in time, but these behaviors may not be pair-wise consistent in any useful sense. However, we insist that the set of behaviors exhibit an overall ***evaluative consistency***. That is, on different occasions, an agent may appear to perform (possibly) conflicting behaviors with respect to the given object, and yet, the overall effect of all the behaviors may still be considered as favorable or unfavorable with respect to the object under consideration. Thus, according to this interpretation, attitudes are evidenced by an overall evaluative consistency. Attitudes may persist even during

periods of behavioral quiescence and are thus generally more persistent and more inclusive than motives. Therefore, the predisposition refers neither to a particular behavior nor to a class of behaviors, but rather to the overall favorability of a behavioral pattern. Consequently, possessing a partial knowledge of an agent's attitude does not necessarily permit prediction of any behavior on its part.

In AI, the term attitude has so far been used to denote concepts such as plans [Pollack 1990], beliefs, goals, intentions, commitments, etc [Rao and Georgeff 1991]. However, we use the term attitude as used in social psychology where an attitude controls the overall mental, physical and communicative behavior of an agent towards an object with the ultimate purpose of either favoring it or disfavoring it (*bi-directional* evaluation). When the object involved is a mental object such as a plan, intention, or commitment, holding an attitude towards it typically requires meta-level reasoning. Attitudes have several characteristics, and we briefly discuss them below.

Persistence and Change

Since one primary use of attitudes is to guide an agent's behavior in *"confusing"* situations (thus helping other coexisting agents to at least partially predict the over all behavior of this agent), attitudes once adopted, must persist for a *reasonable* period of time. This means the agent's mental and physical behavior should be such that any other *"distractions"* the agent might encounter should be either ignored or postponed, while holding onto the current attitude. The set of intentions and commitments that are generated as a result of holding onto the attitude should also include the persistence requirements of the attitude. The extent to which the agent should strive towards holding onto a given attitude in dynamic worlds is included in the behavioral specifications of the attitude. Complex agents, before changing onto new attitudes, may sometimes choose to vary the degree of the currently held attitudes according to new environmental conditions (refer to section 3).

Attitudes and Behavior

When an attitude is adopted, the agent has to exhibit a behavior considered appropriate for that attitude. There are several requirements to this behavior. Firstly, in a dynamic multiagent world, this behavior must include responses to all situations including unexpected state changes, failures of current activities, and changes in other agents' mental and physical behaviors. Secondly, the sub-behaviors of a given behavior must have an overall consistency over the period of time during which the agent is holding that attitude.

Abstract Attitudes

The notion of overall consistency in an agent's attitude based behavior suggests that several attitudes towards a goal will direct the agent to perform behaviors that are non-conflicting and have effects that are consistently favorable or non-favorable to the goal. This concept indicates that several attitudes can be grouped together and leads naturally to the notion of abstract attitude. Two types of abstractions are possible; abstractions over time intervals and abstractions over physical objects and agents. Let $\boldsymbol{K}$ be an attitude towards an object x, denoted $\boldsymbol{K}(x)$. Let $x_1,...,x_n$ be the components of x. The attitude $\boldsymbol{K}$ consists of two parts: $\boldsymbol{K}_1(x_1),...,\boldsymbol{K}_n(x_n)$ where $\boldsymbol{Ki}$'s are (sub)attitudes; and attitude K' is the attitude towards the organizational structure of x. The sub-attitudes Ki's produce the necessary sub-behaviors; and the behavior specified by K' takes into account of the fact that the object is composed of $x_1,...,x_n$. The organizational structure of x typically uses Allen's temporal relations [Allen 1984] for objects such as plans and the spatial relations for physical objects (walls, doors, etc).

Meta Attitudes

When the object x happens to be an attitude, the attitude $\boldsymbol{K}$ towards x becomes a meta-attitude. Meta-attitudes can be used by the agents to generate more controlled behaviors in response to changes in the world. Meta-attitudes are particularly useful during joint activities amongst multiple agents. For example, when an agent A_1 *promises* that it is committing to a plan p, then it is advisable that the agent A_2 attach an attitude value to this *promise* such as $\boldsymbol{confident}_{A2}(\boldsymbol{promise}_{A1}(p))$ which means A_2 is ***confident*** that A_1 will keep its ***promise*** towards p. Attitudes in general, and meta-attitudes in particular, are thus invaluable in multiagent dynamic worlds. The number of levels meta-attitudes can be nested will ultimately depend on the domain of applications. For example, certain forms of beliefs are in fact viewed as attitudes. Thus, beliefs about beliefs, beliefs about beliefs about beliefs, etc. can all be viewed as meta-attitudes. In a joint activity, it is recommended that agents hold mutual beliefs [Levesque 1990]. Mutual beliefs demand infinitely nesting meta-attitudes. The depth of the nesting will not only depend on the application domain, but also on the agent's ability to reason about the depth of nesting.

In addition, during problem solving, it may become necessary for agents to hold more than one attitude towards an object at the same time. Some attitudes may have to be held intermittently; that is, the agent may hold the attitude for an interval T1, drop it at the end of T1, pick it up later for some more time T2, and so on. Thus, the total duration for which the attitude is held is the sum T of T1 and T2, but the attitude itself is not held continuously.

Fire World Attitudes

A fire world is dynamic, typically multiagent, and hostile to the agents. An agent has to respond to the events in the world both reactively and deliberately. Depending on reactivity alone the agent may not be able to solve its problems in the fire world. An important feature of deliberative behavior in a domain like the fire world is that an agent may have to suppress its reactivity as part of its deliberative actions. In such situations, agents can use attitudes to guide their deliberative actions to achieve appropriate behaviors. Attitudes that are useful in a fire world may be classified into several categories, of which the following are some of the important ones: attitudes towards the *objects* in the world, attitudes towards *processes* and attitudes towards *mental objects*.

World Attitudes

An attitude towards a physical object x in the world have the following attributes (see example below).

Example The following describes an attitude of the agent towards an object called *fire1*.

- *Name of the attitude*: ***SmallFire***.
- *Description of object*: **name**: fire1; **description**: fire on chair1 *and* fire on table1.
- *Evaluation*: fire is *small* (multidirectional);
- *Basic agent behavior towards the object fire1*: attempt to put out fire1, and warn other agents.
- *Persistence of the attitude*: if fire1 becomes a medium fire, but the actions in the immediate future segment of the plan includes putting out part of this fire, then the fire will still be considered as small; otherwise, the fire is considered medium or large, in which case this attitude is dropped.
- *Multiple attitudes*: when the fire is closer to chemicals, hold on to the attitude ***SmallFire***, but also adopt another attitude ***dangerous*** (that is, fire is still considered small but also dangerous);
- *Type*: intermittent.

Attitudes towards mental objects requires similarly several attributes. For brevity, we only give the attributes for plan execution as an example, because in our implementation, plan execution is the most dominant activity that an agent performs.

Plan Execution Attitudes

Let p be a plan that the agent wants to execute to reach a destination. This plan can be executed with several attitudes. For example, it may be necessary to execute without interrupting, and some over several attempts, etc. Many interesting attitudes are possible: ***escape-exec(p), normal-exec(p), urgent-exec(p), and execute-quickly(p).*** In multiagent cases, ***cooperation, coordination, team, group***, and so on can be all viewed as joint attitudes.

Problem Solving

In our model, we have used four fundamental concepts to generate the desirable behaviors of the agents: beliefs, attitude, intentions, and commitments. An agent's attitude towards an object is based on the beliefs the agent holds about that object. Attitudes generate appropriate intentions which result in commitments and ultimately action execution. An attitude does not predispose an agent to perform any specific behavior, but rather it leads to a set of intentions that indicate a certain amount of effects toward the object in question. Each of these intentions is related to a specific behavior, and thus the overall effect of an agent's actions on the object corresponds to the attitude toward the object. Once established, an attitude may influence the formation of new beliefs.

Agent Architecture

The agent architecture of our agent has three important modules: a goal generator, a plan structure module (called the Time Line Structure – TLS), and an executor. The goal generator generates goals in response to changes in the world and the messages received. The goals are then placed on the TLS, where the executor will plan for the goal and execute the plan. The strategy used in planning and execution will depend on the *attitude* the agent is holding towards the goal and the derived plan.

Modeling Attitudes

The goal generator generates four types of goals: physical (states of physical objects), communicative, mental and behavioral. Physical goals are goals for physical action such as moving and communicative goals are goals to communicate to others such as screaming. Mental goals specify the internal states (states of TLS in our implementation) to be achieved such as abandoning another goal while behavioral goals specify the behavior to be performed such as escape. Goals of the first three types are achieved by planning and executing the plans. Behavioral goals are achieved by executing the predefined rule bases. An agent can have attitudes towards any object specified in the TLS: goals, plans, actions, and rules (specifying behaviors). We model attitudes by sets of rules which generate goals of one of the four types. In the experiments reported in the next section, we only consider attitudes towards physical goals and mental goals. The plans are placed on the TLS according to some (appropriate for the chosen attitude) temporal relations (one of the thirteen temporal relations discussed in [Allen 1990]).

Performance

We evaluated the performance of the agents that have attitudes built into them in a simulated fire world. The fire world is a large research campus with several buildings and

rooms, containing many objects such as chairs and chemicals. Fire is arbitrarily set and let propagate, and the behavior of the agents were studied over several simulation cycles. Several experiments were performed and the results are summarized below.

Experiment 1: Attitudes

In this experiment, the performance of an agent with different attitudes was studied. The task of the agent was to save property as fire breaks out in multiple places in a room. We consider three types of attitudes: ***quickly***, ***leisurely***, and ***opportunistic***. In the ***quickly*** attitude, the agent attends to the newly generated goal immediately after finishing the current goal. In the ***leisurely*** attitude, the agent attends to the new goal after completing all the goals scheduled on its TLS. The third attitude ***opportunistic*** refers to the behavior where the agent sometimes uses the ***quickly*** attitude and sometimes the ***leisurely*** attitude as the agent thinks is appropriate. Plans were represented progressively abstractly [Au and Parameswaran 1998], and attitudes were used to modify the plan structure in response to world changes.

Fig. 1 show that it is beneficial to have some kind of attitude rather than having no attitudes at all. It is because attitudes provides the agent with the power of meta level reasoning. The performance of the attitudeless agent declines rapidly as the rate of change (of the world) increases, reaching its minimum very quickly (r=8) where it stabilizes. Agents which had adopted attitudes had a larger range of effective operation and showed resistance and tolerance to changes with their performance declining slowly.

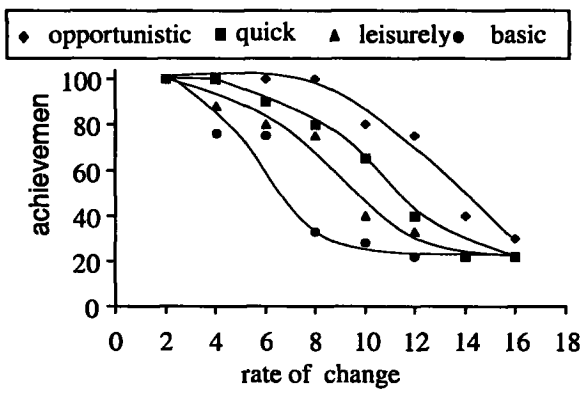

Fig 1. Achievement vs rate of change

Among the three attitudes, the attitude ***quickly*** performs better than the attitude ***leisurely*** because, in the ***quickly*** attitude, the fire was controlled during its initial phase before it became unmanageable. Overall, A_O was the best of the three as it used a meta-attitude taking advantage of both attitudes ***urgently*** and ***leisurely***. Notice that even though all agents performed poorly when the world changed very quickly, the agents were still able achieve a non-zero success rate. This was because no matter how rapid the world changes were, the agents were able to save the first object that caught fire. (We ignored situations when fire occurred and disappeared so fast that the agents were not able to respond to it.)

Experiment 2: Persistence of Attitudes

In Experiment 1, the agent did not make any "*conscious*" effort to hold on to its attitude. As the world changes, agents may need to respond to the changes by dropping one attitude and adopting another. In this experiment, we evaluate the performance of an agent that holds on to its chosen attitude longer than it was expected to, despite the harmful situations the agent encounters in the world.

Holding a particular attitude longer than it was meant to might affect the very survival of the agent. For example, some of the reactive behaviors may be affected when the agent persists too strongly on a chosen attitude, leading to the detriment of its health. Fig 2 shows an interesting behavior exhibited by the agent in the fire world. It shows the effect of sacrificing the agent's health in order to achieve a given goal. Operational range (OR) at a given rate refers to the measure of "freedom" the agent has towards achieving a given goal sacrificing its health but without dying. At the lower rate, the operational range is narrow since to trade off health to achieve its goal, the world itself did not offer much opportunity to risk one's health. Similarly, at higher rates, the range once again is too narrow, since opportunity is less as the world changes too fast. For intermediate values, however, the agent finds more opportunity to risk its health for the sake of achieving the goal. Thus, we find that, at a given rate of change while the risk taking agents have greater possibilities of achieving a given goal and still manage to survive in the world, highly health conscious agents face the risk of not being able to solve the given problem but also the face of risk of being trapped in the world often leading to death ultimately.

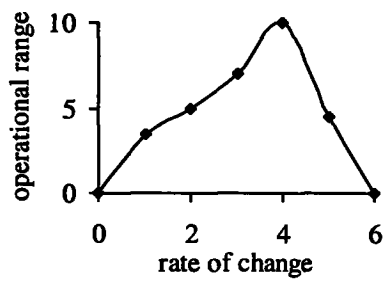

Fig. 2. Operational range Vs rate.

Experiment 3: Meta-attitudes

The final experiment investigates the use of meta-attitudes in problem solving. Clearly, we would prefer an agent that minimizes on its health penalty and but still manages to solve its goal. In this experiment, we introduce an agent which is capable of adopting meta attitudes. Meta attitudes control the degree of base level attitudes by varying their

strength. Meta attitudes were implemented using a (meta level) goal generator which generated meta level goals specifying the level of commitment the agent had to have towards its adopted base level attitudes. Fig. 3 shows that this agent is more tolerant to world changes than the one which did not have this attitude. The dotted lines represent performance of agents holding base level attitudes only, while the solid line depicts the positive effects of the meta attitude.

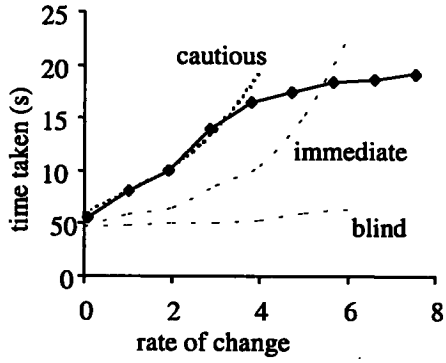

Fig. 3 Meta Attitude Vs. Base Level Attitudes.

Several other experiments were also conducted where attitudes were used to control the behavior of the agents to achieve coordination amongst multiple agents, and in problem solving that spread over prolonged periods of time. To summarize the results, when agents did not have attitudes, they performed poorly (and in fact died too early as the fire spread in many cases). Attitudes gave the agents the necessary mental behaviors which helped them manipulate their internal states in response to changes in the world so that agents could not only achieve their goals, but also exhibited acceptable behaviors during the course of problem solving (such as choosing to stay too close to large fires, walk over fire, etc. when no other options existed to save its life).

Related Work and Conclusion

In AI, the term attitude has been used to denote different concepts. For example, Pollack[Pollack 90] refers to plans as mental attitudes. [Kalenka and Jennings 1995] identified ***responsibility***, ***helpfulness***, and ***cooperativeness*** as three important social attitudes that may prevail in social problem solving. It is also common among the researchers to refer to beliefs, goals, and intentions as attitudes. In this paper, we have viewed attitude towards an object x as a mechanism which generated an appropriate meta-level behavior with regard to that object. We also argued that attitudes can be complex as abstract attitudes, meta attitudes, and joint attitudes. Intention does not always say how soon a goal must be achieved, and commitment does not say how strong it should be. The role of attitudes is to specify this extra parameter so that agents can adapt their behaviours to new situations.

Georgeff et al. [Georgeff et al 1998] report experiments similar to our Experiment 1 where the agent attempts to maximize its score using intention as a guide to filter currently available options. In our case, the behavior is much more complex where depending on the attitudes adopted, the agent exhibits different types of mental behaviors.

In dynamic worlds, agents are not only required to strive for solving their goals, but are also expected to exhibit only acceptable behaviors during the process of problem solving. Attitudes are necessary to guarantee this behaviour while at the same time maintaining consistency and predictability of behaviors across multiple agents when they are involved in collective problem solving. Attitudes guide not only the individual agent's behaviors, but also the behaviour of a group of agents. We are currently investigating application of attitudes to collective problem solving and the results are reported in [Madhu 2001].

References

[Allen 1984] J. Allen. Towards a General Theory of Action and Time. AI Magazine, Vol.23 1984. Pp 123-154.

[Au and Parameswaran 1998] Au S., N. Parameswaran, Progressive Plan Execution in a Dynamic World, Planning Workshop in AIPS 1998.

[Dict 1992] The Angus & Robertson Dictionary and Thesaurus, HarperCollins Publishers Sydney 2001, 1992.

[Allport 1935] Allport G W Attitudes in C M Murchison (ed) Handbook of Social Psychology, Worcester, Mass, Clark Univ Press, 1935.

[Fishbein 1975] Fishbein M and Ajzen I Belief, Attitude, Intention, and Behaviour *An Introduction to Theory and Research* , Addison-Wesley Publishing Company, 1975.

[Kalenka and Jennings 1995] S. Kalenka and N.R. Jennings: On Social Attitudes: A Preliminary Report, International Workshop on Decentralized Intelligent and Multi-Agent Systems, Krakow, Poland, Nov. 22-24, 1995.

[Levesque 1990] Hector Levesque, Jose Nunes and Philip Cohen. On Acting Together. National Conference on Artificial Intelligence, 1990.

[Georgeff et al 1998] M. Georgeff, B. Pell, M. Pollack, M. Tambe and M. Wooldridge. The Belief-Desire-Intention Model of Agency. In Proceeding of the 5th International Workshop, ATAL 1998.

[Madhu 2001] PhD Thesis: Team Activities in Multiagent World. UNSW 2000.

[Pollack 1990] M. E. Pollack, "Plans as Complex Mental Attitudes," in P. R. Cohen, J. Morgan, and M. E. Pollack, eds., Intentions in Communication, MIT Press, Cambridge, MA, 1990.

[Rao and Georgeff 1991] A. Rao and M. Georgeff. Modeling rational agents within a DBI-architecture. KR&R 1991, pp 473-484.