From: AAAI Technical Report SS-95-02. Compilation copyright © 1995, AAAI (www.aaai.org). All rights reserved.

A Distributed Adaptive Control Architecture for Autonomous Agents

Bruce L. Digney
Canadian Department of National Defence, DCIEM Robotics Laboratory,
1133 Sheppard Ave West, North York, Ontario, M3M 3B9, CANADA,
e_mail digney@dciem.dnd.ca

Abstract

Recently considerable interest in behavior-based robots has been generated by industrial, space and defence related activities. Such independent robots are envisioned to perform tasks where safety or economic factors prevent direct human control and communication difficulties prevent easy remote control. Although many successes have been reported using behavior-based robots with prespecified skills and behaviors, it is clear that there are many more applications where learning and adaptation are required. In this research, a method whereby reinforcement learning can be combined into a behavior based control system is presented. Skills and behaviors which are impossible or impractical to embed as predetermined responses are learned by the robot through exploration and discovery using a temporal difference reinforcement learning technique. This results in what is referred to as a Distributed Adaptive Control System (DACS), in effect the robot's artificial nervous system. Presented in this paper is only a general overview of the DACS architecture with many details neglected. A DACS is then developed for a simulated quadruped mobile robot. The locomotion and body coordination behavioral levels are isolated and evaluated.

1 Introduction

Although conventional control and artificial intelligence researchers have made many advances, neither approach seems capable of realizing autonomous operation. That is, neither one can produce machines which can interact with the world with an ease comparable to that of humans or at least higher animals. In response to such limitations, many researchers have looked to biologically/physiologically based systems as the motivation to design artificial systems [1]. As an example are the behavior-based systems of Brooks [2] and Beer [3]. Although many variations exist behavior-based control systems generally consist of an interacting structure of simple behavior modules. In some cases each module is responsible for the sensory motor responses of a particular behavioral level. The overall effect is that higher level behaviors are recursively built upon lower ones and the resulting system operates in a self-organizing manner. Both Brooks' and Beer's systems were loosely based upon the nervous systems of insects. These artificial insects operated with a set of prespecified and hardwired (non-adapting) rules and were shown to exhibit an interesting repertoire of simple behaviors. Although this approach has shown successes, it is obvious that there are many situations where predetermined solutions are impossible or impractical to obtain. It is therefore proposed that by incorporating learning into the behavior-based control system, these difficult behaviors could be acquired autonomously through exploration, discovery and learning.

Complex behaviors are usually characterized by a sequence of actions with a critical error signal (success or failure) with success becoming evident only at the end of that sequence. Thus, the required learning mechanism must be capable of both reinforcement learning as well as temporal credit assignment. Incremental dynamic programming techniques such as Barto and Sutton's [4] temporal difference (TD) appear to be well suited to such tasks. Based upon their previous work in adaptive heuristic critics [5], TD employs adaptive state and action evaluation functions to incrementally improve an action policy until successful operation is attained. The incorporation of TD learning into behavior based control results in a framework of Adaptive Behavior Modules (ABMs) which is referred to here as a Distributed Adaptive Control System (DACS). This paper provides only a brief description of the DACS and ABMs and the implementation of the interacting behavioral levels of locomotion and body coordination on a simulated quadruped mobile robot. A more detailed development is available in elsewhere[6]. Examples of other behavioral levels for the robot, including global navigation, task planning and task coordination are implemented and discussed by Digney [7] [8]. Additional work,

concerned with improving DACS viability in real applications for autonomous, semi-autonomous, tele-operated and tele-autonomous operations, is currently being pursued.

Although this work was performed in the context of mobile robotics the concepts developed can be applied to areas of process control and intelligent manufacturing. Manufacturing plants with such adaptive nervous systems may be able learn to operate efficiently and then adapt as to changes that occur. Furthermore, by increasing the intelligent capabilities of the plant the human effort required during product transitions would be greatly reduced as the plant itself would be in part responsible of the learning and configuring to the new tasks.

2 Adaptive Behavior Modules

Within a behavior-based control system, each individual behavioral level is established through connections to the environment via sensors and to actuators or to other behaviors via command signals. The individual behavioral levels of a DACS are established in a similar manner. Sensory connections are used to establish the system state, goal state, and the sensory based reinforcement vectors. Command and reinforcement connections are used to connect individual ABMs within the DACS's framework. The system state vector represents the connections to sensors which establish the system state of the behavioral level. The goal state vector represents the connections to sensors which define system state locations or transitions which may be of use to higher behavioral levels. The sensory based reinforcement vector represents the connections to sensors monitoring the condition of the robot's components that might be damaged by the actions or inactions of that behavioral level.

It is acknowledged that these connections do represent a hand decomposition into levels of behavior and as such do constitute a form of predetermined knowledge. However, in most applications these connections are straight forward and it is the internal control strategies composing the skills and behaviors which are dependent upon unforeseeable and changing conditions and are often impossible or impractical to predetermine. It is not until the robot's configuration and capabilities allows for more general operation and that its desired tasks and environment types require the autonomous generation of behavioral levels, that emergent and adaptive connections become necessary. As robots and their tasks become more general purpose (more animal like) that emergent structures must be addressed. The concepts of emergent and adaptive DACS structures and how they inter-relate with the mechanisms of simulated evolution, autonomous discovery and learning, external conditioning, communication and collective intelligence are discussed elsewhere [7]. A generic adaptive behavior module is shown in Figure 1(a). Details of the command, reinforcement and sensory signals and connections as well as the details of the internal learning and adaptive mechanisms of the adaptive behavior module can be found elsewhere [6] [7].

3 Distributed Adaptive Control Systems

Using the ABM as the adaptive behavioral level building block, a distributed adaptive control system can be developed. Figure 1(b) schematically illustrates the DACS as being loosely divided into a hierarchy of interacting behavioral levels. At the highest behavior level the desired goals are specified with an externally supplied goal based reinforcement signal. Whenever the actions, or more likely a sequence of actions, of the actuators have successfully performed the desired task a favorable goal based reinforcement results at the highest behavioral level. Otherwise, an unfavorable goal based reinforcement signal is present. At the lowest level are the actuators which interact with the environment and ultimately perform the high level goals. It is the responsibility of the DACS to learn all the intermediate skills and behaviors required to bridge the gap between the desired goal and the actuators. These intermediate behavior levels are combined in a recursive manner to form increasingly more complex behaviors. Such complex behaviors cannot be realized using a single adaptive sensory-response coupling, but require interaction between many behavioral levels.

Figure 1(b) also shows the conceptual reinforcement signal flow directions. Environmental based reinforcement originates at the actuators and flows upward toward the highest behavioral level. Sensory based reinforcement originates at all behavioral levels and also flows upward towards the highest behavioral level. Goal based reinforcement can be thought of as originating at the highest behavior level and flowing downward throughout all of the branches of the DACS. These three signals then compete to influence the operating characteristics of each individual behavioral level and effectively the collective characteristics of the robot.

Changes and malfunctions can confront the robot and affect the DACS at any behavioral level. To what extent the adaptive framework is affected depends upon the severity of the change. Two categories of change are defined: severe and non-severe. If adaptation of a single behavioral level is adequate for recovery a non-severe change is said to have occurred. If the adaptation of adjacent behavioral levels higher up in the framework is required for recovery, then a severe change is said to have occurred.

It is straight forward to extend the DACS beyond an individual and to encompass populations of robots. The collective and cooperative behavioral levels that result would learn to control groups of robots rather than groups of actuators. In the absence of a higher collective behavior commanding an individual robot, the robot

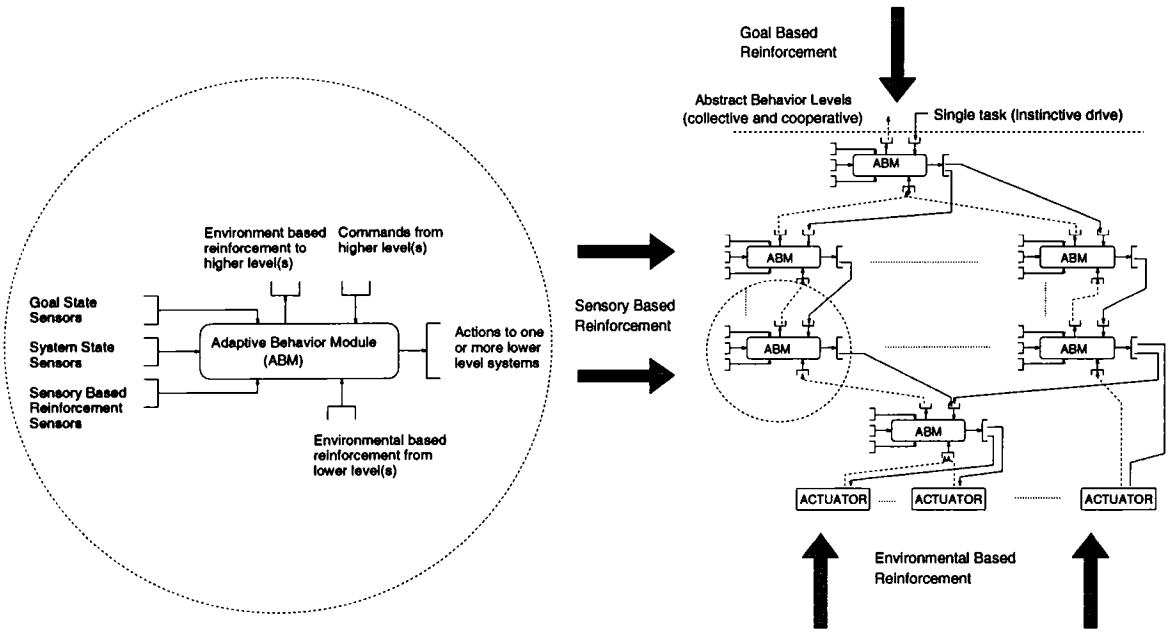

Figure 1: Left (a): Adaptive Behavior Module. In generic form, the behavioral level is established by its sensory connections to the world as well as by its command and reinforcement connections within the DACS's adaptive framework. Right (b): Schematic of Typical DACS. Each reinforcement signal's point of origin and flow direction is indicated by the large arrows.

can be made to learn some desired task by defining an instinctive drive. The instinctive drive of a robot is specified as some system state transition at the highest behavioral level of the DACS that establishes the purpose of the robot. This system state transition is externally specified as being desirable and results in a favorable goal based reinforcement signal. This single external reinforcement signal contains no information from which a correct response or specific skills and behaviors can be inferred. This signal can only be used to drive the robot's DACS to discover and learn the skills and behaviors necessary to fulfill this instinctive drive.

4 DACS for a Quadruped Mobile Robot

4.1 Robot and DACS Configuration

A simulated quadruped mobile robot was used to investigate the DACS architecture. The robot was equipped with all the actuator and sensory systems required for successful operation in its intended environment. It had four, two degree of freedom legs and a hopper into which it could load and transport substances. If controlled properly, these actuators would be sufficient for the performance of its intended tasks as well as for its survival. The robot was capable of perceiving its external world using visual, sonar and tactile sensors as well as monitoring its own internal states. At the chosen level of abstraction, the robot model neglected dynamics and the sensors and actuators were assumed ideal, although in other work [7] a tolerance to non-ideal sensors and actuators was demonstrated.

A DACS was developed for the robot which was then given an open-ended task embedded as an instinctive drive. The robot was externally rewarded whenever it eradicated or rendered benign an undesirable substance from its world. Once the robot was placed inside the simulated three dimensional world it was left to develop skills and behaviors as it interacted with its environment.

The DACS used to control the robot is shown schematically in Figure 2(a). It consists of six adaptive behavior modules distributed over the behavioral levels of locomotion (LM), body coordination (BC), local navigation (LN), global navigation (GN), task planning (TP) and task coordination (TC). The robot's actuators for its leg and loading mechanisms are shown at the bottom of that hierarchy. Figure 2(a) shows only the command and reinforcement signal paths between behavioral levels. In this paper only the two interacting behavioral levels of locomotion and body coordination will be investigated.

4.2 Locomotion (LM)

Although not the most efficient method of locomotion, walking provides many interesting and challenging problems for the proposed learning and adaptive systems. Quadrupedal walking requires the learning of complex actuator sequences in the midst of numerous false goal state locations and modes of failure. Quadruped loco-

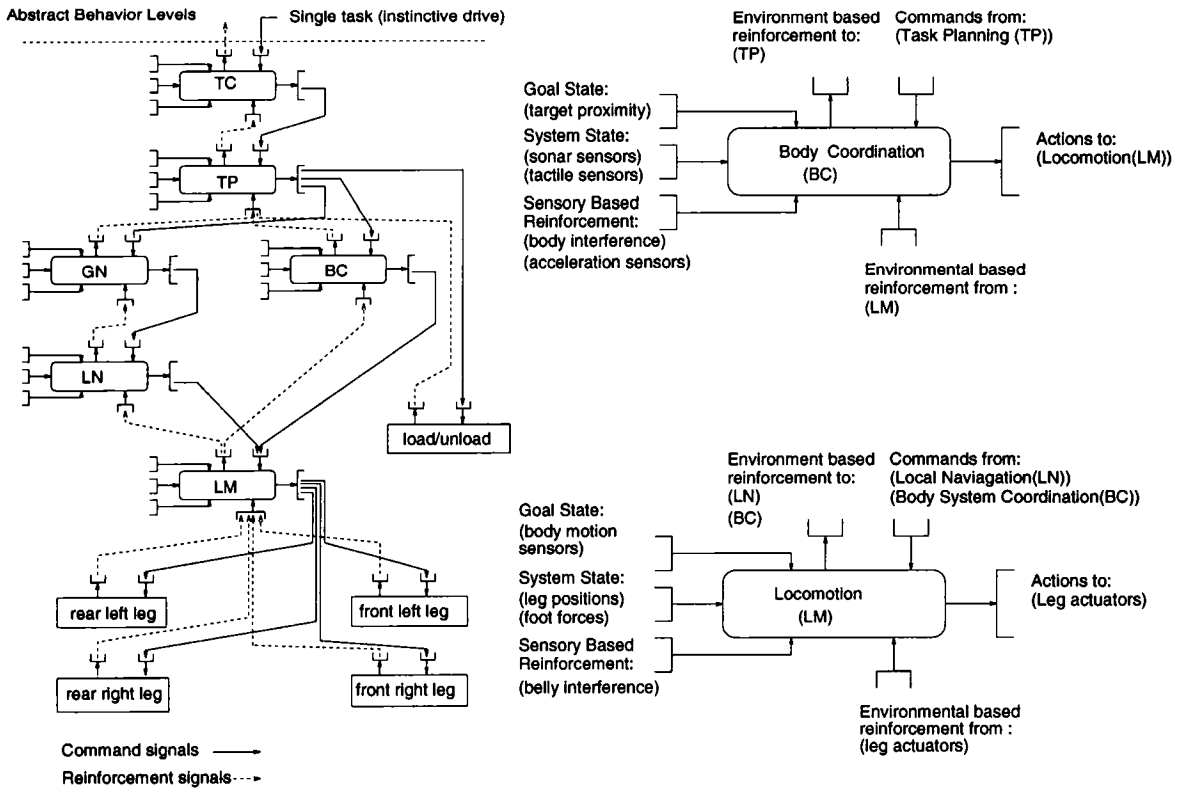

Figure 2: Left(a): DACS for a Single Quadruped Robot. In (a) the dashed lines represent environmental based reinforcement signals and the solid lines represent command signals. Right(b):Locomotion and Body Coordination Behavioral Levels. In (b) the adaptive behavior module for learning locomotion and body coordination skills for the mobile robot are detailed with the sensory, reinforcement and command connections as indicated.

motion particularly exploits the temporal credit assignment and cyclic goal capabilities of the DACS. Shown in Figure 2(b) is the locomotion ABM with its sensory, reinforcement and motor action connections.

4.3 Body Coordination (BC)

The body coordination ABM utilizes the goal states of the lower level locomotion ABM. The gaits of the locomotion ABM become valid commands which are to be used by the body coordination ABM. These movements are referenced relative to some goal substance location and are often performed in the vicinity of local features. When a substance of interest enters the robot's local sensor range it is the responsibility of the body coordination ABM to coordinate the body movements in such a way as to discover and learn useful behaviors. This behavioral level is established in general form by its goal state vector, but the exact behaviors that are possible are unknown. They are expected to include moving the robot into loading position or moving it away from the target substance. The body coordination ABM with its sensory, reinforcement and action connections is shown in Figure 2(b). Body coordination is an intermediate behavior level which does not interact with the world directly, but indirectly through the locomotion ABM. The complexity of the skills capable of being learned in this adaptive hierarchical structure becomes evident at this behavioral level.

5 Simulation and Results

5.1 Locomotion

5.1.1 Physical Constraints and Assumptions

The responsibility of the locomotion ABM was to discover all possible gaits and then learn the actuator sequences required to perform them. Quadruped locomotion over a solid surface involved both balance and movement in the desired direction. The robot was assumed to be balanced, with its body free of contact with the surface of the world whenever a straight line could be drawn between any two legs in contact with the ground that passed through the robot's center of gravity. In effect, at least two diagonally opposite legs had to be in contact with the ground at the same time for the robot to be balanced. In this simulation, the quadruped mobile robot was assumed to be moving if all the legs in

contact with the ground were applying forces such that their collective effort resulted in the desired motion of the robot body. If the desired motion resulted a 0 valued goal base reinforcement signal replaced the otherwise high negative value of R_g. No movement occurred if any leg was applying a force that opposed the desired motion and a reinforcement of R_{high} was then sent from that actuator. R_{nom} and R_{high} reinforcements resulted from the normal and restricted or over-extended leg actuator movements, respectively. Sensory based reinforcement was used to condition the robot to walk in a safe manner. When operating on a solid surface, the safety of the robot required that the robot move with its body free of the ground. Hence, any resulting gaits which the robot learned had to also suspend and balance the robot above the ground. The body interference sensor on the belly of the robot supplied R_{high} sensory based reinforcement when it was in contact with a damaging surface and 0 whenever the robot was safely suspended.

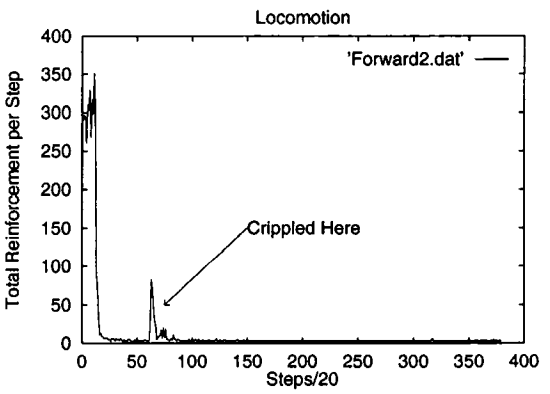

Figure 3: The Performance of the Locomotion ABM. The performance of the locomotion behavioral level while learning the forward gait. Note the initial period of poor performance and the rapid improvement to optimum gaits. The recovery from an actuator malfunction is also indicated.

5.1.2 Results

The locomotion ABM quickly discovered and learned its possible skills. Initially, the leg movements were random as the robot began exploring both its own internal operating characteristics and its external surroundings. As the locomotion ABM gained experience, four realizable gaits emerged. After the robot had mastered all of its realizable gaits, an internal malfunction was introduced. This malfunction specifically involved the disabling of the horizontal actuator on a single leg. Effectively, the robot was then capable of only supporting itself and incapable of applying any horizontal forces with that disabled leg. After a short period of adaptation the robot recovered and had re-learned all the gaits and transitions. If the malfunction had rendered the robot incapable of a previously learned gait, adaptation of higher behavioral level(s) would have been required and a severe change would have occurred.

Figure 3 shows the total reinforcement signal for the forward gait experienced during the initial learning and recovery an malfunction. The vertical axis of this performance curve shows the total reinforcement occurring for each successful step taken and the horizontal axis indicates the number of steps taken. In this plot the robot was alternating between gaits at ten step intervals. This allowed for transitions between gaits to be learned and investigated as well. Gait learning was evident in the rapid improvement section of Figure 3, while the section of slower improvement was where the gait transitions were learned. The converged leg sequence for the intact forward gaits is shown in Figure 4. Considering the end of the robot closest to the reader as the front, leg_0, leg_1, leg_2 and leg_3 are the right front, left front, left rear, and right rear legs, respectively. The actuator extensions, v_{ex} and h_{ex}, resulted in downward and forward movements of a leg, respectively. The actuator retractions, v_{rt} and h_{rt}, resulted in upward and backward movements of the leg, respectively. The intact forward gait's command, c^{leg_0}, c^{leg_1}, c^{leg_2} and c^{leg_3}, and reinforcement signals, r_e, r_g and r_s, are listed in Table 1. Environmental, goal and sensory based reinforcement signals are represented by r_e, r_g and r_s, respectively.

5.2 Body Coordination

5.2.1 Physical Constraints and Assumptions

Once a substance was detected within range of the local sensor system, it became the responsibility of the body coordination ABM to coordinate the robot's actions such that useful behaviors resulted. These behaviors included the learning of movement sequences to either capture or avoid the target substance. When a desired behavior was completed a 0 valued goal based reinforcement resulted while a high negative value of R_g was present otherwise. In this particular environment contact of the robot body with a rock was damaging and undesirable. Any such contact was detected or felt by the body interference sensors and resulted in a R_{high} sensory based reinforcement signal. Furthermore, the presence of a rock physically prevented further motion in that direction. Contact with a hole was also considered undesirable and resulted in R_{high} sensory based reinforcement from the internal acceleration sensors.

5.2.2 Results

When considering the two possible behaviors of avoid and capture, it was seen that the capture behavior required the learning of the most challenging and interest-

Table 1: Command and Reinforcement Signals: *Final Forward Gait Control Strategy.*

Frame	Actions				Reinforcements			Comments
k	c^{leg_0}	c^{leg_1}	c^{leg_2}	c^{leg_3}	r_e	r_g	r_s	
1	v_{rt}	v_{ex}	v_{rt}	v_{ex}	R_{nom}	R_g	0	No forward motion (leg reset motion)
2	h_{ex}	h_{rt}	h_{ex}	h_{rt}	R_{nom}	0	0	Forward motion (productive motion)
3	v_{ex}	v_{rt}	v_{ex}	v_{rt}	R_{nom}	R_g	0	No forward motion (leg reset motion)
4	h_{rt}	h_{ex}	h_{rt}	h_{ex}	R_{nom}	0	0	Forward motion (productive motion)

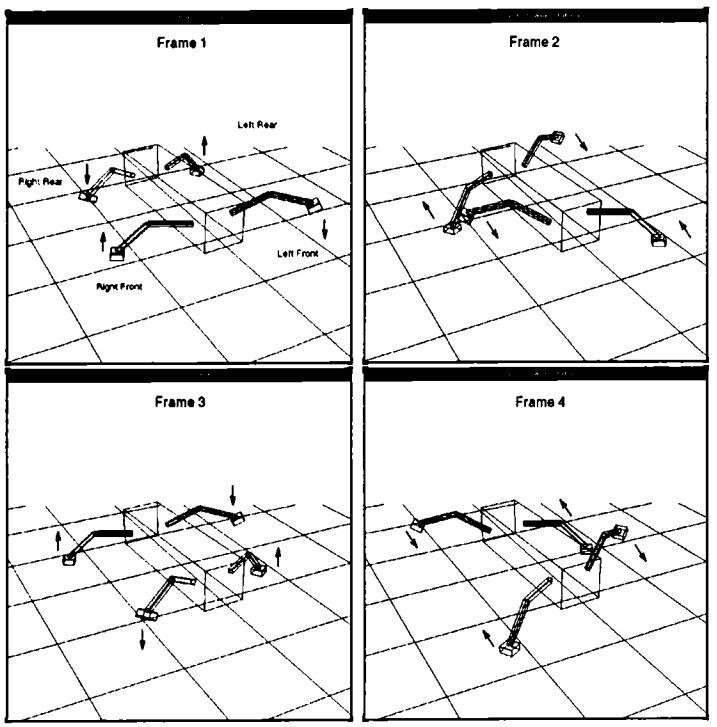

Figure 4: Leg Sequence of Locomotion ABM: *Forward.* Final leg actuator sequence for the forward gait control strategy. Note the action sequence is, frame 1, 2, 3, 4, 1, 2, 3, $\cdots$.

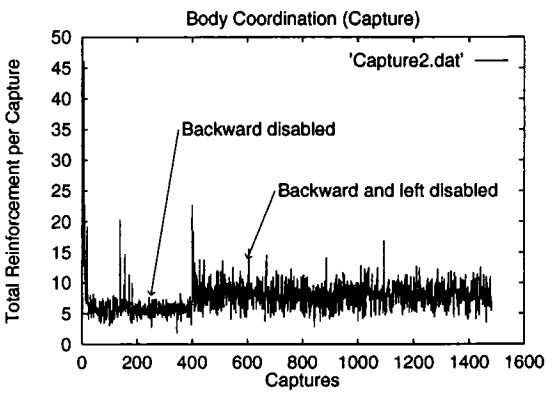

Figure 5: The Performance of the Body Coordination ABM. The performance of the body coordination behavioral level while learning the capture behavior. Note the period of poor performance eventually improving to some optimum level and the recovery from the disabling of the left turn gait.

ing movement strategies, while the avoid behavior simply resulted in the robot scurrying away from the target. Therefore, the capture behavior was chosen to be examined in detail. To make this behavior even more difficult, the robot was considered permanently incapable of performing the backward gait.

The performance of the robot at learning the capture behavior is shown in Figure 5. The vertical axis shows the total reinforcement encountered per target capture and the horizontal axis indicates the current capture attempt. This plot shows initial poor performance which eventually improved to some optimum performance. In addition to having the backward gait disabled, the robot then had its left turn gait disabled. The body coordination ABM then relearned the capture behavior with only the forward and right turn gaits available. Initially the robot wandered around unproductively, hitting rocks and going through holes. Eventually a safe, efficient and productive capture behavior was learned as indicated in Figure 6 and Table 2.

6 Conclusions

The research described in this paper was initiated to realize a higher level of autonomy in robotic control systems. This resulted in an intelligent robot capable of autonomously learning tasks bounded by its sensory and physical capabilities. This increase in autonomy was attained by the incorporation of physiologically motivated learning and adaptation techniques within the robot's control system rather than by simply increasing the amount of predetermined and embedded information. The robot learned what it needed during operation and was not burdened by, or limited to the predetermined actions and responses of its designers. It is clear that the level of abstraction used in these simulations is a significant departure from reality. However, these simplifications were necessary for these preliminary investigations. Ongoing work in continuous state and actions spaces, emergent structures, robot conditioning, robot-

Table 2: Command and Reinforcement Signals: *Final Capture Control Strategy.*

Step k	Actions c^{LM}	Reinforcements r_e	r_g	r_s	Comments
1	*left*	R_{nom}	R_g	0	Turns in line with target (as detected by sonar sensors)
2	*forward*	R_{nom}	R_g	0	Contacts hole with (as detected by tactile sensors)
3	*right*	R_{nom}	R_g	0	Turns to avoid hole (as detected by tactile sensors)
4	*forward*	R_{nom}	R_g	0	Goes around hole (as detected by tactile sensors)
5	*left*	R_{nom}	R_g	0	In line with target (as detected by sonar sensors)
5	*forward*	R_{nom}	R_g	0	Target close (as detected by sonar sensors)
6	*forward*	R_{nom}	0	0	Target captured (as detected by proximity sensor)

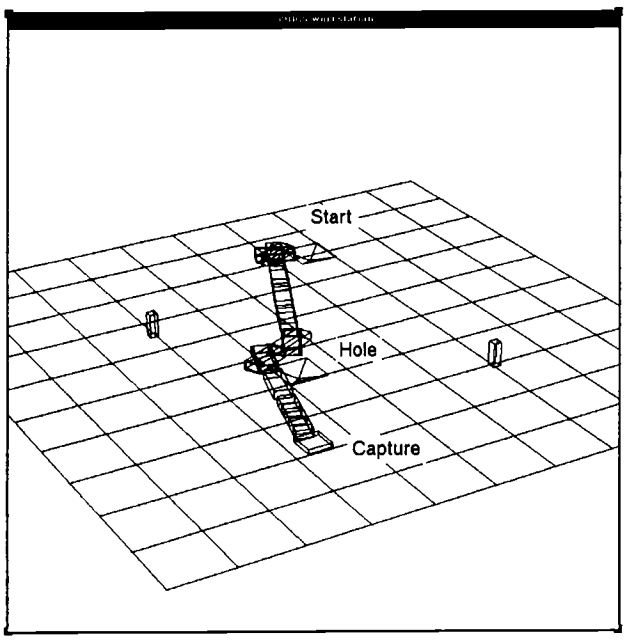

Figure 6: The quadruped mobile robot using a final converged control strategy to capture the substance. Note the avoidance of obstacles and the efficient performance. The rocks are represented by the tall blocks, the target by the flat plate, holes by the inverted pyramids and the robot by the sequence of long blocks.

human interfaces and tele-autonomy is leading to implementation on a real robot operating in a real environment.

7 Acknowledgments

Part of this work was performed while the author was with the Intelligent Systems Research Laboratory and the Department of Mechanical Engineering at the University of Saskatchewan, Saskatoon, Saskatchewan, CANADA.

References

[1] Meyer, J.A. and Guillot A. (1994), From SAB90 to SAB94, *From animals to animats 3: The third conference on the Simulation of Adaptive Behavior SAB 94*, Brighton UK, August 1994, pp 2-11, MIT Press-Bradford Books, Massachussets.

[2] Brooks, R. (1986), A Robust Layered Control System For A Mobile Robot, *IEEE Journal of Robotics and Automation*, Vol. 2, no. 1, pp. 14-23.

[3] Beer, R.D., Chiel, H.J., and Sterling, L.S., (1990), A Biological Perspective on Autonomous Agent Design, *Robotics and Autonomous Systems*, Vol. 6, pp. 169-186.

[4] Barto, A.G., Sutton R.S. and Watkins C.H. (1989), Learning and Sequential Decision Making, *COINS Technical Report.*

[5] Barto, A.G., Sutton, R.S., and Anderson, C.W. (1983), Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems, *IEEE Transactions on Systems, Man, and Cybernetics SMC-13*, pp. 834-846.

[6] Digney B. L., (1994) A Distributed Adaptive Control System for a Quadruped Mobile Robot, *From animals to animats 3: The third conference on the Simulation of Adaptive Behavior SAB 94*, Brighton UK, August 1994, pp 344-354, MIT Press-Bradford Books, Massachussets.

[7] Digney, B.L. (1994) Emergent Intelligence in A Distributed Adaptive Control System. *Ph.D. Thesis*, University of Saskatchewan, Saskatoon, Saskatchewan, CANADA.

[8] Digney B. L., (1993) A Distributed Adaptive Control System for a Quadruped Mobile Robot, *IEEE ICNN International Conference on Neural Networks*, San Francisco, March 1993, pp. 144-150.