Learning Mobility: Adaptive control algorithms for the Novel Unmanned Ground Vehicle (NUGV)[1]

Mike Blackburn, SSC San Diego
(619) 553-1904, mike@spawar.navy.mil

1. Summary

Mobility is a serious limiting factor in the usefulness of unmanned ground vehicles. This paper contains a description of our approach to develop control algorithms for the Novel Unmanned Ground Vehicle (NUGV) to address this problem. The NUGV is a six-degree-of-freedom, sensor-rich small mobile robot designed to demonstrate auto-learning capabilities for the improvement of mobility through variegated terrain. The learning processes we plan to implement are composed of classical and operant conditionings of novel responses built upon pre-defined fixed action patterns. The fixed action patterns will be in turn modulated by pre-defined low-level reactive behaviors that, as unconditioned responses, should continuously serve to maintain the viability of the robot during the activations of the fixed action patterns and of the higher-order (conditioned) behaviors. The sensors of the internal environment that govern the low-level reactive behaviors also serve as the criteria for operant conditioning. Using this adaptive controller, the NUGV should learn to negotiate difficult obstacles, and to protect itself from collisions and falls.

[1] At the time of this writing, the NUGV is in the final stages of detail design and prototyping by Automated Controlled Environments, Inc., 25133 Avenue Tibbitts, Unit A, Valencia, CA 91355, (661) 775-7754 Fax: (661) 775-7770, under contract N66001-02-M-X105, with support from the Office of the Secretary of Defense Joint Robotics Program (JRP). The author gratefully appreciates the support of the JRP Coordinator, and the assistance of ACEi in the preparation of this manuscript.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **22 MAR 2003**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Learning Mobility: Adaptive Control Aglorithms for the Novel Unmanned Ground Vehicle (NUGV)**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Mike /Blackburn**		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Space and Naval Warfare Systems Center 53406 Woodward Road San Diego, CA 92152**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**		
13. SUPPLEMENTARY NOTES **The original document contains color images.**		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**53**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Contents

Figures

Tables

2. Objectives

2.1. To What Should We Aspire?

The standard to which we should aspire in the control processes of our robots is not an indolent, ineffectual, and operator's attention consuming automaton, but rather a mobile, self-sufficient, loyal, cooperative and obedient agent, somewhat like a hundred-pound Golden Retriever.

2.2. Why Should We Aspire to This?

We conceive of our robot as an aid to the operator, not the other way around. Thus our robot should be there when the operator needs it, ready to assist. Otherwise, the robot should stay out of the way, and take care of itself.

2.3. Is This Not Science Fiction?

This will not be science fiction if we define carefully the needs of the robot and install low-level control process on the robot to provide for these needs. Second, if we couple one or more of the solutions to the critical needs of the robot with some activity of its operators, the probability that the robot will track, trail, and learn to cooperate with its human operators should be increased.

2.4. Resolving the Conflict

The reader may sense a contradiction here. I suggest above that the robot must have low-level control processes that permit it to take care of itself, while at the same time state that these must be coupled to activities of the human operator so that the robot is in some way dependent upon that operator. Can we have both independence and dependence, or self-interest coincident with social-interest? Can the robot exercise independence by virtue of its low-level control processes, and then become dependent upon a human operator through the acquisition of higher order robot behaviors that also provide service to the operator? In the following, I will attempt to explain how we can. Resolving this conflict between self-interest and social-interest should provide for the usefulness of the robot[2].

---

[2] Many of the terms that I will use in this discussion come from biology and psychology. They are thus loaded with anthropomorphic connotations. I hope that the reader does not become too suspicious at this point, but looks for my later description of algorithms that will implement these concepts in the artificial system of robot hardware and software.

3. Control with Independent Agents

To achieve our objectives, we will implement a control architecture that differs significantly from the principal approach taken today in mobile robotics. First, we view our robot as an independent agent, and will attempt to endow it with all of the necessary capabilities to promote its own welfare. Second, we view our own role in the process more as director and collaborator, than as user and operator, and, as such, will employ methods of control that involve more the leash than the lever.

3.1. Control by Negation

To the degree that the robot will be self-controlled it will also be self-motivated. Then, as it is self-motivated, the operator[3] may be excluded from giving explicit instructions on the direction and intensity of any robot action. Rather, the director (or human collaborator) should be able to provide information on the intended objective, to which the robot would then be socially motivated to pursue. The director may then observe the progress of the robot toward that objective, and intervene only as necessary to veto or negate a particular action that the robot is attempting to execute. Once an action is negated, the robot would, on its own initiative, select a different approach to the objective[4].

3.2. The Purpose of Local Control is Preservation

An agent is useful only while it is viable. An agent's viability is preserved when it remains physically intact, its sensors and effectors function as designed, and its energy reserves are adequate for any exigencies. Factors that jeopardize these conditions are variously extremes of temperature, shock and other collisions, and un-replenished power consumption.

3.3. Homeostasis is an Optimal State for Preservation

Homeostasis is the state of the agent that optimally predisposes it to perform some additional activities within its present environment[5]. Thus an agent preserves itself by performing activities that maintain its homeostasis and by avoiding actions that seriously disturb its homeostasis. Various internal sensors measure the state of the agent, defining its homeostasis. At a very low level of control, these sensors are coupled with subsystems that enable the activity of the agent. When a subsystem is failing, the agent's activity is threatened, and some change in activity should occur to restore the subsystem functionality, in other words, to restore homeostasis.

[3] The term operator is inconsistent with the control of an independent agent. We operate machines, and political operatives are defined by their ability to operate politicians, but we direct actors and employees.
[4] Negation will be effective only to the degree that the operator can both tempt and threaten the robot, and to the degree that the robot can generate alternative actions to achieve the intended objective. We will discuss these possibilities later in this paper.
[5] The permission of additional behavior is also known as *survival*.

4. *A Control Architecture for Independence and Survival*

The control processes (algorithms) for our robot must execute within the constraints imposed upon it by our mechanical design, sensors, electronics, and a few (very few) behavioral preferences. All of these things we provide to the robot in assembly, and are analogous to the ontological implementation of a genetic code.

4.1. The Physical Constraints

Developing a local control capability through the use of artificial intelligence (AI) algorithms should prove feasible in an embodied system such as our Novel UGV. The physical system of the Novel UGV provides not only constraints, but also a means to complete feedback loops with the environment that is essential for stability.

The physical equipment of our robot, that will enable and constrain our AI algorithms, is shown in Figures 1, 2 and 3. The Novel UGV is composed of three principal segments, a central core, and two pods. All three segments contain electrical power, power transmission mechanisms, sensors for both the internal and external environments, radios for inter-pod communication, and electronics for local processing. The core contains radios for communication with the operator control unit (OCU).

The two pods are tracked for conventional tank-type motion across planar surfaces. The pods are each connected to the central core by a single axle, about which the pods can rotate. These two axles are mounted at either end of the core, and laterally near the end of each pod. The axles, with pods attached, can rotate about the ends of the core.

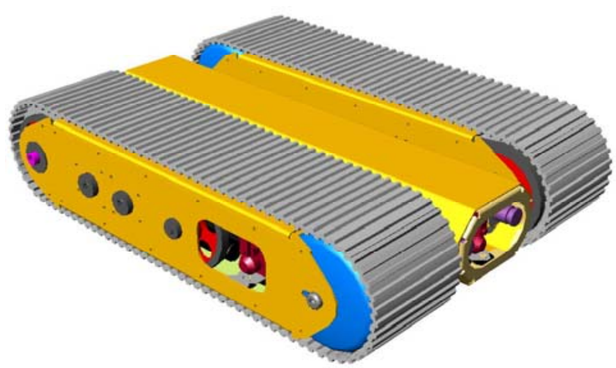

Figure 1. Outer appearance of the Novel UGV

The NUGV is symmetrical on all major axes, so that if the image in Figure 1 was rotated 180 degrees in any direction, it would appear the same. Sensors for the external environment (video cameras, SONAR, and IR proximity detectors) are located on both ends of the core faceplates, and (sans video cameras) on the outboard sides of the two pods.

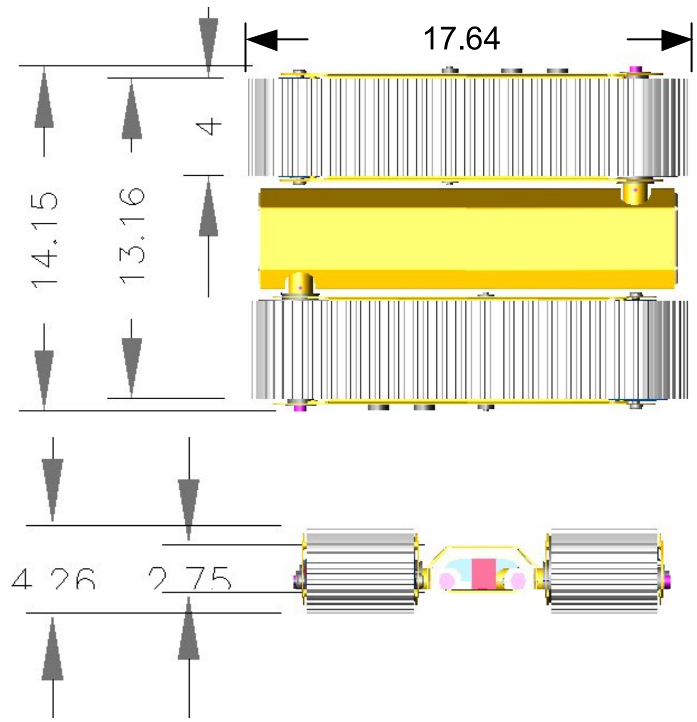

Figure 2. Exterior dimensions of the Novel UGV in inches.

The total weight of the first prototype Novel UGV should be approximately 30 pounds. The use of lighter materials in its construction should reduce this weight by about 30%. The vehicle may scale upwards to increase payload and energy storage capacities. Downward scalability will be limited by the availability of suitably scaled electronics, energy transmissions, sensors, and energy density storage or recovery devices. Recent developments in micro-electromechanical systems (MEMS) promise to significantly push back limitations to the first three, but micro energy storage or recovery issues are yet to be addressed.

Figure 3. Cutaway of the Novel UGV showing components.

The layout of some of the internal components of the NUGV can be seen in Figure 3. Batteries are represented by gray cylinders, circuit board are shown in light green, and power transmission devices are shown as black cylinders and belts. Many wires, cables, and other obscuring components are not shown for clarity.

4.2. Multiple Degrees of Motion Freedom Allow Multiple Conformations

The physical architecture of our robot permits it to assume several different conformations. Our physical architecture enables six mobility degrees of freedom[6]. For comparison, the Foster-Miller *Talon* tracked robot has two, the iRobot *PackBot* tracked robot with flipper assist has three, and the Sony *SDR-4X* humanoid robot has twenty-eight, more or less. A sample of the different conformations that are possible with the Novel UGV's six degrees of freedom is given in Figure 4. The variable conformation of the vehicle permits a large diversity of behavioral responses to environmental conditions. In general the degree of behavioral complexity possible in a mobile agent is a non-linear function of the mobility degrees of freedom.

Each of the conformations depicted in Figure 4 can be achieved or passed through by a variety of combinations of pod motions.

[6] The simultaneous remote control of six mobility degrees of freedom would pose a significant challenge for a human operator. For this and other reasons we intend to automate much of the local control processes. As we expand the number of mobility degrees of freedom in order to increase the opportunity for increased behavioral complexity in our development of even more capable robots, this local automation will take on even greater importance.

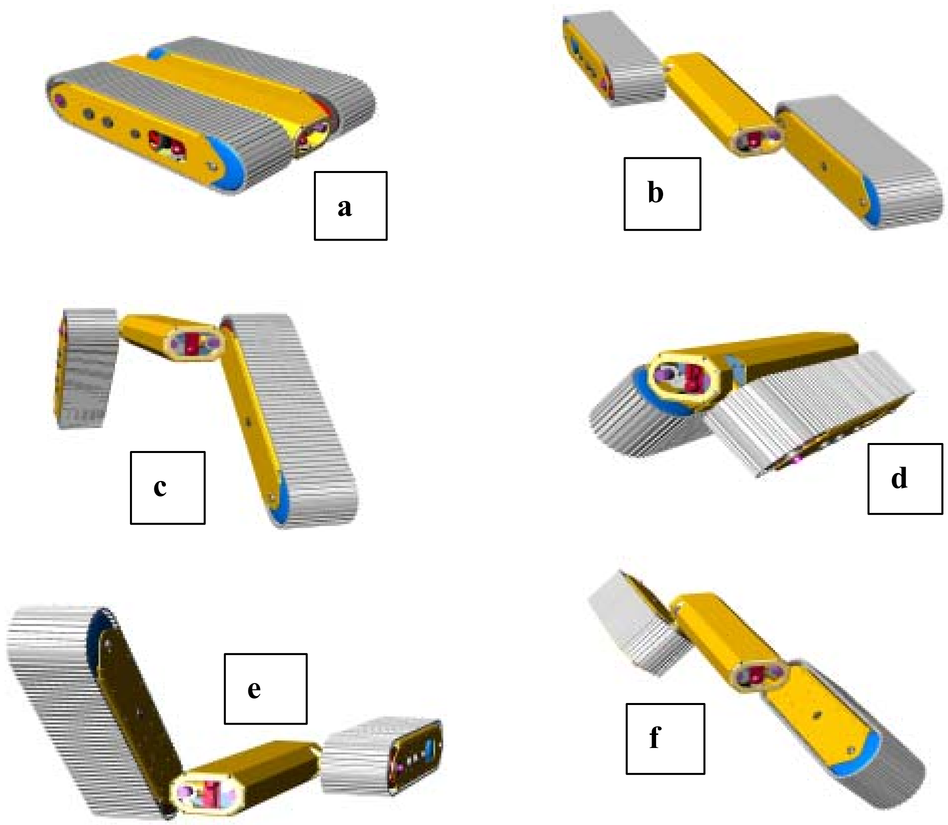

Figure 4. Various possible robot conformations.

Of the robot's six degrees of freedom of movement, the two degrees of freedom associated with the camber axes are limited in transit, while the other four degrees of freedom are rotationally continuous.

Each conformation shown in Figure 4 will have a different utility for one of the different topologies of the surfaces over which the robot will attempt to move. Because the robot is symmetrical along its lateral (X, side to side), coronal (Y, top to bottom), and sagittal (Z, end to end) axes, there will always be two absolute conformations with respect to gravity that will accomplish the same task in the same way.

Given a planar surface with small physical texture relative to the vehicle, the most efficient conformation of the robot is expected to be that of Figure 4.a. The vehicle is most stable in this conformation as the maximum amount of track contact with the planar surface is possible and the vehicle has the lowest center of gravity. From this conformation the vehicle could execute turns by skid steering wherein the track velocities are varied between the pods to rotate the vehicle while in place or while progressing.

The conformation shown in Figure 4.b, the open position, could be most useful when a high barrier must be scaled, or when a narrow chasm or gulf (negative obstacle) with a width not in excess of the length of one pod must be crossed.

The conformation in Figure 4.c could be useful for elevating the cameras for improved perspective, and for passing over occasional obstacle clumps.

The conformations of Figures 4.d and 4.f could permit the vehicle to maintain stable traction on irregular surfaces such as beams, tree branches, gabled rooftops, and pipes (inside or out). This conformation would also permit the vehicle to avoid high centering on boulders and other irregularities in the plane of traversal.

The conformation in Figure 4.e represents the pose the robot might take in approaching a step change on a planar surface.

The choice of conformations for any set of environmental conditions would have to depend upon the robot's ability to assess those conditions, and recall previous conformations that accomplished a task objective and met the optimization criteria.

A second problem is the morphing from one conformation to the more optimal conformation without losing friction or balance. I will address these problems progressively through the paper.

4.3. Information Flow During Control

The different components of the physical architecture fall within the following classes:

- Sensors of the internal environment
- Sensors of the external environment
- Effectors composed of motors and transmission elements
- Energy storage composed of batteries
- Computational resources

The computational resources provide the substrate for connectivity matrices between sensors and effectors. Theses matrices are composed of fixed and plastic elements

These components are graphically shown in Figure 5. The arrows of Figure 5 indicate the direction of information flow. The control laws are embedded in the two boxes labeled "fixed connections" and "plastic connections". The fixed connections are established primarily by design, while the plastic connections are established primarily through the vehicle's experience in operation, though based upon pre-defined mechanisms. Feedback is indicated in the horizontal arrows between the boxes of connections, and in the line through the environment that provides information on the physical consequences of the robot's behavior.

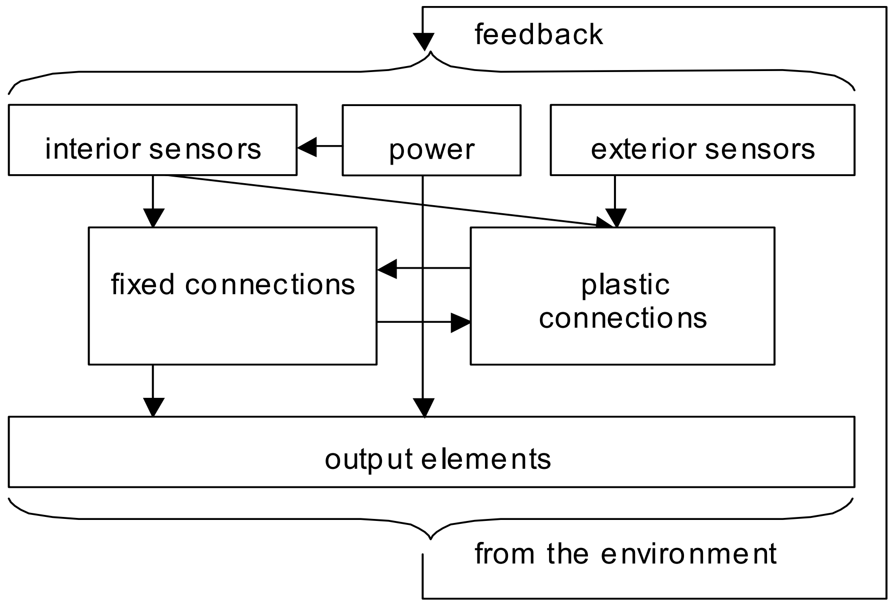

Figure 5. Schematic Architecture of the Novel UGV Control System

4.4. Representation of the Control Output

Since our Novel UGV is symmetrical on all axes, we can define top vs. bottom, left vs. right, and front vs. back only with respect to gravity and to the direction in which the vehicle is moving. The six motors therefore can have an absolute identification and a relative identification. For most of our discussion I will use the relative identification, recognizing that the core sensors for gravity and direction of motion will have to route the motor commands (M) to the appropriate motors in the appropriate way to execute the desired action.

Each of our six motors can turn in either direction. We represent this by 12 output elements. The torque on the motors will be proportional to applied voltage. We represent the applied voltage by the numerical value on the output element. Thus we have the following elements in our motor vector (**M**)[7]:

CL, for camber left pod,
CLx, for camber left pod counter clockwise,
CR, for camber right pod,
CRx, for camber right pod counter clockwise,
RL, for rotation of left pod,
RLx, for rotation of left pod counter clockwise,
RR, for rotation of right pod,

[7] Throughout this paper, I will indicate vector variables by bold type, and scalar variables in regular type.

RRx, for rotation of right pod counter clockwise,
TL, for track rotation of left pod,
TLx, for track rotation of left pod counter clockwise,
TR, track rotation of right pod, and
TRx, track rotation of right pod counter clockwise.

Where x is always a counter clockwise rotation from the perspective of the vehicle. In general, to get the track pods coordinated in the direction of travel of the vehicle, one pod must move clockwise while the other pod moves counter clockwise.

As it is impossible for any one of the motors to turn in both directions at the same time, we should provide for contradictory commands to the same motor to cancel at the output element. For example:

$$M_{CL} = CL - CLx$$

4.5. Sensors of the Internal Environment

We have a sensor field composed of numerous sensors of the internal environment. These include nine accelerometers (three for each of two pods, and three for the core), three core magnetometers, four track rotation sensors (two per pod), sixteen touch sensitive whiskers (four on the ends of each of the pods), two core faceplate pressure sensors (one at each end of the core), eighty plate pressure sensors, three battery voltage sensors (one in each compartment), and three battery current sensors.

The pod plate pressure sensors, the touch-sensitive whiskers, and the core faceplate pressure sensors would not ordinarily be considered sensors for the internal environment, but we include them here because they basically require physical contact with an external object to produce an output. Thus they are neither predictive of contact, nor descriptive of the typology of the immediate environment.

A vector of features, derived from the nine accelerometers, defines the conformation of the vehicle (**C**). By measuring the acceleration vector in each trio of accelerometers and comparing the measurements to each other, the relationship (tilt and camber) of the pods to the core can be determined. When the motors are all quiescent, gravity is the only influence on the accelerometers, and the accelerometer input is sufficient for an unambiguous determination of conformation. Some examples of this vector are shown in Figure 6.

When the pods are in motion with respect to the core, the pod accelerometers will sense both the pod motion acceleration and gravity. The two effects will be confounded. The conformation will be changing during these motion-imposed accelerations. The effects of motion acceleration could be extracted from the effects of gravity as there is a very predictable effect on the accelerometers with the different changes in axle position due to activations of the motors. However, we yet do not have sensors for axle rotation. Thus there will remain some conformation uncertainty until the pod and camber rotations stop.

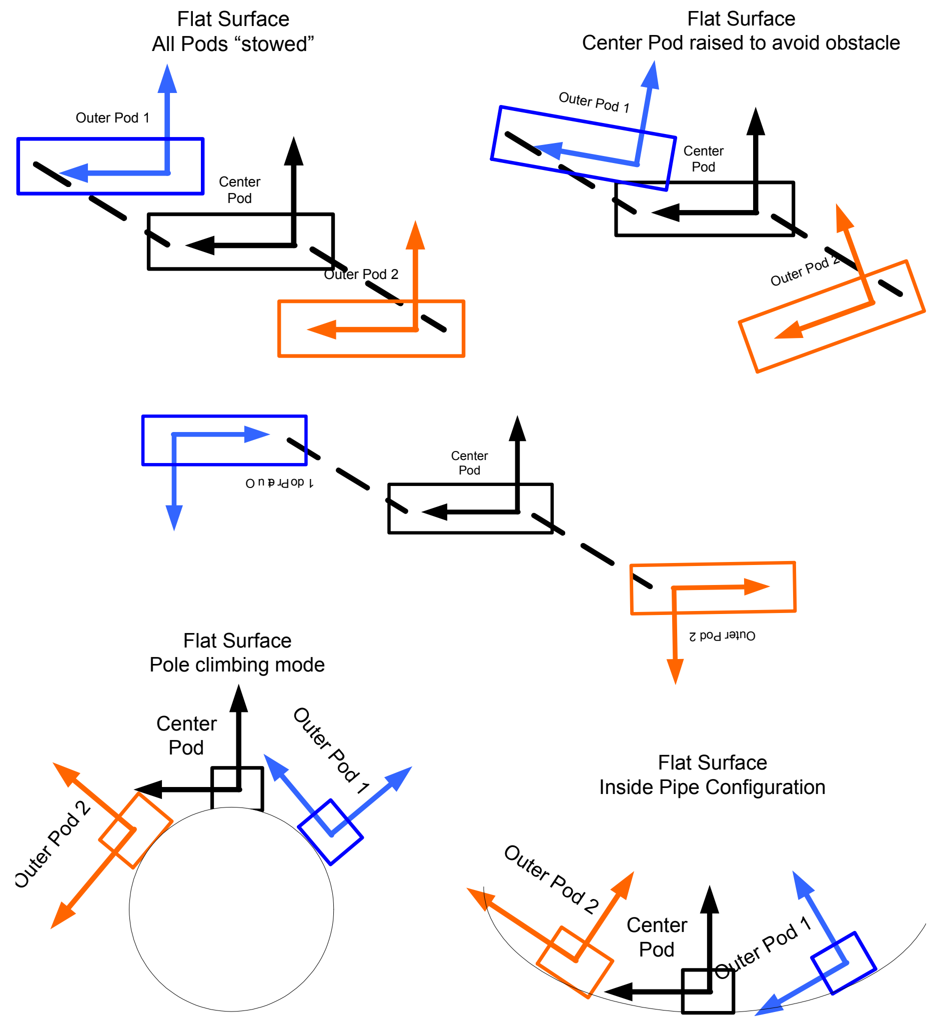

Figure 6. Accelerometer indicators of conformation.

The robot would compare information from its internal sensors to determine its present conformation, and to assess the success of any attempts to change its conformation, but the control algorithms have available information from all sensors at all times, some of which may be irrelevant to the particular control decision, in this case – conformation, but which later may become a disambiguifying factor. For example, during changes in conformation, the pod plate pressure sensors and the whiskers will cooperate with the accelerometers to determine whether the robot’s contacts are due either to the ground plane, to an obstacle, or to an appropriate leverage point.

4.6. Fixed Action Patterns (FAP)

To control the six degrees of freedom during translation, and during the transition from one conformation to the next, the robot will likely need several different behaviors composed of sets of coordinated motor commands. Similar organized behaviors, specific to the physical makeup of an animal, and stereotypical in nature, are called Fixed Action Patterns (FAP) in the Neuroethology community. I will use that term here as well. The robot's Fixed Action Patterns exhibit a predictable set of events characterized by coordinated motor torques and timing. The Fixed Action Patterns do not necessarily depend upon any particular environmental conditions, but may be invoked by triggers related to the above sensors of the internal environment.

A network of delay elements that can be invoked as a unit will define each of the six FAP. The connectivity of the elements in those units will define the sequence and strength of commands to the 12 output elements. The sub-networks that manage the different FAP are located in the box labeled *fixed connections* in Figure 5. The FAP progress by the strength of the recent history of current pattern to evoke the next element of the pattern. Thus, baring any changes in the external and internal environments, a pattern, once initiated, may continue in an infinite loop. The impossibility of an infinite behavioral loop, however, is obvious, as behavior itself will produce changes in both environments, disrupting the behavior.

4.6.1. Fixed Action Pattern P. Porpoising

FAP-P may be attempted when the robot is fully immersed in a liquid medium and is neutrally buoyant. Immersion would be sensed with the present sensor suite by the absence of contact information from any of the whiskers or plate pressure sensors. Under these conditions, the rearward track pod would assume a position 180 degrees to the rear and oscillate, while the forward track pod would maintain its normal position with respect to the core and then oscillate in counter phase with the rearward track pod. The net result of the oscillations of the two pods should be a porpoising of the robot through the liquid medium. Diving and surfacing could be accomplished by varying the angles of the forward and rearward pods around which the oscillations are made.

4.6.2. Fixed Action Pattern R. Resting to Running

FAP-R permits the robot to run consistently and rapidly in a particular direction on a smooth planar surface. This FAP prefers the conformation shown in Figure 4.a. Sensor conditions that would favor this FAP are significant pod plate pressure, and the absence of whisker contact. To achieve this conformation, the robot assesses its core accelerometer values. If the Y-axis (see Figure 7) is at +/- 1, the core is horizontal on whatever surface the robot is resting. The robot then attempts to match the Y-axes of the two pods with the core Y-axis value by rotating the pods away from their contact points without upsetting the Y-axis value. A stable surface would permit this maneuver and the robot could then close to its normal preferred conformation.

To execute a run command from the normal closed position, the simplest mechanism would be to have the two track motors run essentially at the same speed. Speed control may be modulated by accelerometers on the X and Z-axes.

Changes in direction of travel could reactively occur in response to the asymmetric detection of obstacles along the trajectory. The sensor detecting the obstacle, activation of pod whiskers for example, could trigger a turn away from the obstacle by biasing the X-axis accelerometer output to the motor controllers. The robot could turn most efficiently by lifting the ends of one or both pods off of the surface during the turn. Which end is lifted could depend upon the desired direction of the turn.

If the robot was positively or negatively (but not neutrally buoyant) the robot could use FAP-R to swim on the surface of a liquid medium or crawl on the bottom of its container respectively.

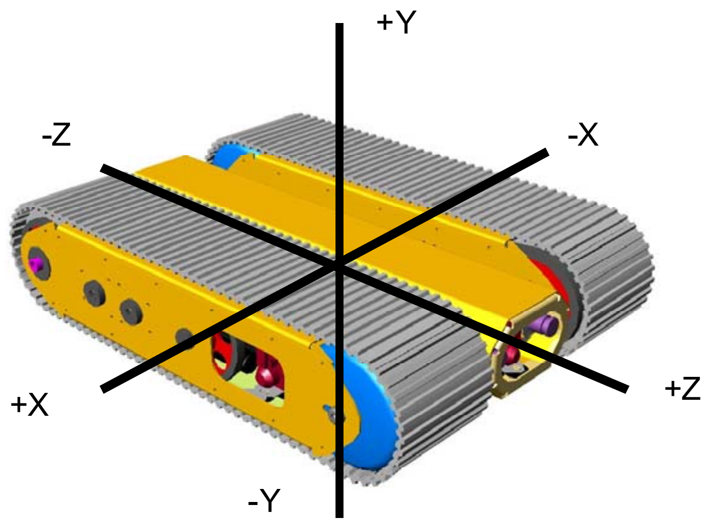

Figure 7. Vehicle Motion Axes

4.6.3. Fixed Action Pattern S. Scaling

FAP-S permits the robot to scale a large non-vertical obstacle by a combination of walking and running. The robot would normally initiate the FAP-S by encountering an obstacle with its whisker sensors. To accomplish scaling, the robot could rotate its two pods outward from the normal closed conformation until contact is reestablished on the pressure plates. If the forward pod is on the right of the vehicle, the rotation of both pods would be counter-clockwise, while the reverse direction of rotation would be performed if the vehicle was inverted at the time of first contact. During rotation, the forward tracked pod would normally make contact with the obstacle before the rearward pod again made contact with the ground plane, and the robot would pull itself up on the obstacle using a combination of its track tread rotations and forward track pod rotation. If the obstacle was short, the rotation could continue and the robot would pull itself over the obstacle. Any unevenness of the obstacle, such as a staircase, could cause the forward pod to continue in its rotation (as it is yet leading and yet in the open conformation so that the same direction of rotation would be maintained, causing the pod to complete a full rotation), and the rearward pod to oscillate or porpoise as the pod attempts to improve its traction with the obstacle.

The robot could perform a descent down a slope by maintaining the same pattern as was used in the ascent.

A similar combination of walking and porpoising could also be used to propel the robot across the surface of a liquid in which it was positively buoyant.

4.6.4. Fixed Action Pattern T. Tumbling

FAP-T permits the robot to tumble by alternately rotating the pods around the core in a consistent direction. One use for tumbling could be to dismount from a straddle position on a beam. A conceivable trigger for this FAP could be the absence of forward and rearward motions by any other FAP. The tumbling could be performed most efficiently from the normal closed position (Figure 4.a). To initiate tumbling, one pod on the side to which tumbling would progress would begin a rotation under the core. After a lag, the second pod would begin its rotation under the core. This would tend to bring the core over the pod with the first rotation. Next, after completing its range of rotation, the direction of rotation would change on the first pod, while the second pod would continue with its rotation progress under the core while the core was being lifted away from the first pod. Upon completion of its rotation transit, the second pod would also reverse its direction of rotation and move to complete the inversion of the platform. As either pod reached the limits of rotation in either direction it would change direction and repeat the process. In this way, the tumbling could be completed. Alternatively, the rates of rotation could differ, with the pod moving faster initially in the direction of the tumble. The rate as well as the direction of rotation could alternate at each range limit. The pods could also rotate on their connecting arms to facilitate tumbling by moving the center of gravity further away from the core.

4.6.5. Fixed Action Pattern U. Undulating.

FAP-U permits the robot to elevate its core above the terrain without moving forward. A conceivable trigger for this behavioral pattern could be the detection of low battery capacity. An elevated core might make the robot easier to find. Other triggers could include loss of RF signal, and SONAR indication of a blocked visual field. Thus elevating the core could also improve radio communications, and it could give the robot's video cameras a better perspective above ground rubble. Accelerometers would provide the primary sensor input during the execution of this FAP. Undulation could begin from the normal closed position by rotating both pods outward. Undulation could proceed by continuing the rotation until the core ascends to its apogee and begins again to descend. The undulation could be halted at this point whereupon the core would be at its most elevated position with respect to the ground plane. Because of the wide tracks, the robot should be stable in this position, but during movement, stability could be achieved either by adjusting the rotation of either pod or by adjusting the direction and rate of the pod track rotations, or both. Continuing the undulation would involve a reversal of the pod rotations at this point. At the point of co-pod rotation where core elevation no longer changes, the direction of pod rotation would again change lifting the core again to its apogee.

From the core perigee, continuing the pod rotations in the same direction would restore the robot conformation to the normal closed position.

To achieve an extended position, useful on steep slopes, the rotation could be interrupted as the core begins to lift from the surface during pod rotation.

4.6.6. Fixed Action Pattern W. Walking

FAP-W permits the robot to walk consistently in a particular direction on a variegated[8] planar surface. In this pattern, the track treads could remain still or continue in rotation, while the pods rotate on the core connection arm in alternating and parallel motions in the direction of travel. FAP-W could evolve as both pods encounter obstacles[9]. The pod whose core connection arm is located at the forward end of the core, as defined by the direction of travel, begins to rotate first. This could be detected by the contact sensors on the pods or on the core faceplate, or by the accelerometer data. The forward pod would rotate forward as in the FAP-S. However, when the first pod was rotated fully forward, the second pod rotation would begin also in the forward direction. This would tend to elevate the core. Afterwards, the two pods could continue with their rotations at equivalent rates, remaining about 180 degrees out of phase, undulating the core up and down over the variegated surface. Turning on such a surface could be accomplished by activating the tracks in addition to the pod rotations, by differentially rotating the pods, and by changing the camber angle of the pods.

4.6.7. Fixed Action Pattern Y. Yawing

FAP-Y may permit the robot to squeeze through a narrow passageway. The trigger for this maneuver could be activation only of the forward outboard pod whiskers while the robot was in the normal closed position. That pattern of activation could indicate a gap through which the robot could attempt to squeeze[10]. The minimum gap width that the present NGV could negotiate is approximately eight inches. This pattern begins by the NUGV backing up and extending the pods outward as in FAP-U, however, at the point where the pods are horizontal with the core, a camber command is triggered that draws both pods in (down with respect to gravity). This maneuver will force the pods to rest on

---

[8] An example of a variegated planar surface over which it would be appropriate for this NUGV to walk would be an extended egg-carton with compartments in sufficient quantity and size to hold 144 basket balls.

[9] The difference between FAP-W and FAP-S could be in the accelerometer indications of core and pod position when the obstacles are encountered. A consistently inclined core indicates the predominance of the obstacle in the forward direction, while an oscillating core indicates a variegated surface that may be better managed by the FAP-W.

[10] Normally FAP-Y would not be attempted when alternative action patterns were available.

the outboard edges of their track treads. Then, alternately rotating further the ends of the pods while moving forward should cause the vehicle to Yaw back and forth. If the rate of Yaw is correct, the vehicle should pass through an orifice of dimension down to the minimum. The principle here is that the pods are alternately rotated while the pod camber angle directs the angle of attack of the treads to turn the vehicle. But a similar pattern may be accomplished by simple skid steering of the vehicle while in the open position.

4.7. Summary of the Fixed Action Patterns

The various Fixed Action Patterns are summarized in Table 1.

Fixed Action Pattern	Trigger	Expected Conditions
P Porpoise	Absence of any contact	Immersion
R Run	Movement commanded by the activity monitor in the absence of whisker output	Obstacle free
S Scale	Obstacle is detected in the forward direction of travel by whiskers. Core accelerometer indicates consistent ascending or descending pattern.	Obstacles
T Tumble	Both forward and reverse motion are blocked	Entrapment
U Undulate	Low battery voltage; obstacle detection; loss of RF input	Poor visibility, poor RF communications, low power reserves.
W Walk	Velocity < expected, obstacles. Core accelerometer indicates inconsistent ascending or descending pattern.	Variegated surface Mud and other impediments
Y Yaw	Outboard whisker activation	Presence of a traversable gap

Table 1. Fixed Action Patterns

The Fixed Action Patterns are low-level behavioral repertoires by which the robot coordinates its movements. The set of Fixed Action Patterns pretty much is inclusive of all of the maneuvers possible with the six degrees of motion freedom that are available to the robot. A greater diversity of overt behavior could be observed when the internal conditions evolve during behavior and trigger transitions among the patterns. These transitions could occur at any time during a behavior, and do not require the completion of one pattern before the initiation of another. The Fixed Action Patterns are behaviors to which the robot would default under certain circumstances that required some behavior but for which requirements for no other task-specific actions were evident.

4.8. Direction and Extent of Pod Rotations Define the Patterns

Five of the seven Fixed Action Patterns (less FAP-T and FAP-Y) differ primarily on the directions and extents of the pod rotation with respect to the core. These differences are shown in Figure 8.

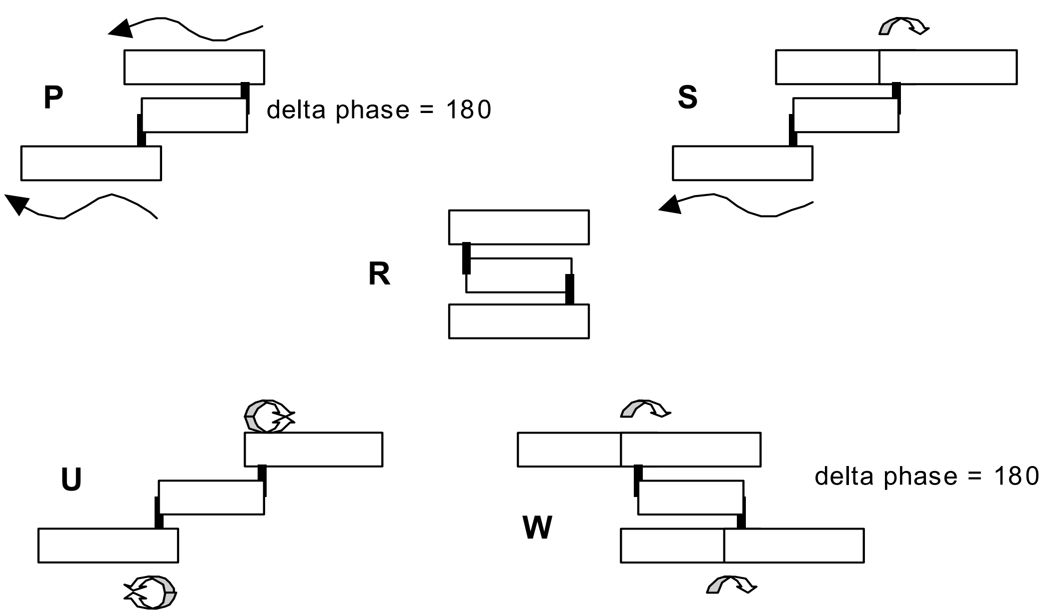

Figure 8. Pod Rotation Differences between five of the seven Fixed Action Patterns

4.9. The Behavioral Constraints

The behavioral constraints are simple reactive behaviors that constrain other behaviors to prevent serious disturbances to homeostasis. Thus I will call these reactive behaviors *Basic Reactive Patterns (BRP)*. The robot will come from the factory equipped with a few pre-planned[11] BRP that respond to critical events in ways that would restore the sensors of those events to their states before the events occurred. The sensors involved are those that monitor the key homeostatic conditions. The BRP occur when certain pre-established sensor threshold values are breached. The sub-networks that manage the

[11] Pre-planned in the sense that the rules that govern the definition of the transfer functions between input and output are pre-determined in the design of the controller, and yet are subject to rapid as well as slow adaptations to improve performance and compensate for hardware drift.

different BRP are located among the boxes labeled *fixed connections* and *plastic connections* in Figure 5.

For our robot, these critical events should be: loss of mobility or inactivity, loss of core balance, loss of track contact, collision with the core face plates, and loss of energy. We selected these critical events for their relevance to the viability of the robot under the conditions with which we expect it to normally operate. Other critical events could be considered, and appropriate sensors supplied, such as for temperature, water infiltration, and tampering, either physically or electronically.

The Basic Reactive Patterns do not necessarily take the robot to any specific location or in any specific direction, but do continually act to correct or shape any ongoing random, pre-programmed (such as the fixed action patterns), and/or acquired behaviors of the robot.

Previously I described seven Fixed Action Patterns that can be assumed by the robot in response to various internal sensor conditions. All of these behaviors are stereotypical in the sense that they are completely predictable given the constellation of sensor conditions in the internal environment. The five Basic Reactive Patterns that I will now describe will constantly modulate the seven Fixed Action Patterns.

4.9.1. Basic Reactive Pattern A. Activity

The objective of BRP-A is to prevent the robot from either moving too slowly or moving too rapidly. Movement may be assessed by the integration of the accelerometers and sensors monitoring the rotations of the track drive wheels. There will be a range of activity that is optimal for the performance of the robot and director. No activity is, by definition, undesirable for a mobile robot. High levels of activity, while potentially useful under extreme circumstances, will more quickly deplete the energy reserves of the vehicle, subject it to destructive collisions, and reduce the usefulness of sensor information that is returned to the human observer during monitoring. Thus the extremes of inactivity and activity should be avoided. To accomplish this, the very low or very high activity readings should contribute to increases or decreases in activity as appropriate to maintain activity within the preferred range.

The Gaussian curve in Figure 9 shows the expected relationship of activity levels and system performance. To optimize performance, the system attempts to keep activity in the preferred mid range by modulating the activity of the 12 output elements. An alternative approach is to use the activity gain to modify the general inhibitory or excitatory influences within the controllers. Specific inhibitory or excitatory control commands needed to execute any particular behavior could be adjusted by these gains. Should we take this approach to modulation, then a slow stealthy movement of the robot could be performed while the accelerometer input was attempting to move the system to the right in Figure 9. Under that circumstance, a sudden decision to execute an evasive maneuver would be facilitated by the elevated excitatory gains and depressed inhibitory gains.

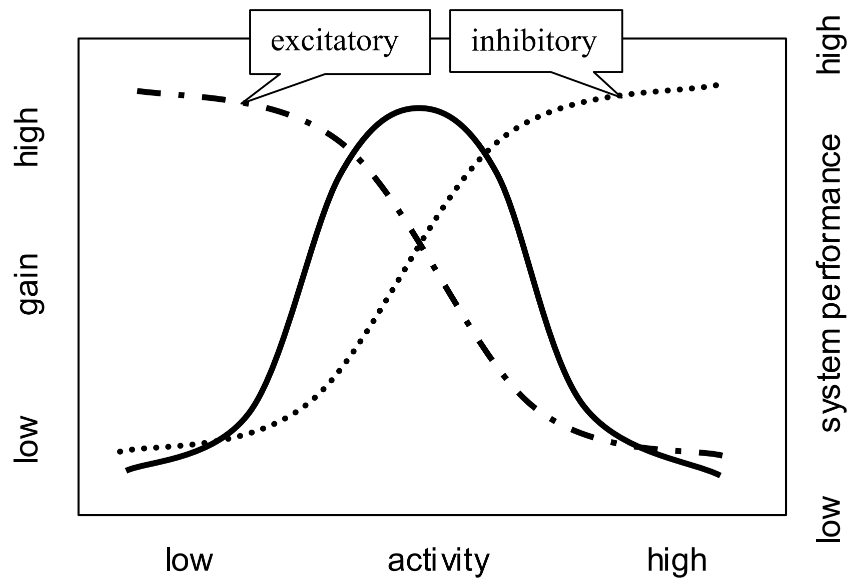

Figure 9. Relationship between preferred activity levels and performance.

In an artificial system, to achieve some independence from the fickle motivations of its operator, the robot must provide an internal provocation that is linked to activity itself. This internal provocation should contribute to an apparent spontaneity that permits trial and error learning and the exercise of learned behavioral patterns.

To move without an explicit or external provocation the robot could have in its control algorithm a parameter that assesses the total dynamics of its actuators. The dynamics of the system are characterized by the accelerometer (A) activity. Let this quantity be D.

Then

$$D_{in} = \Sigma\, \mathbf{A}.$$

D should persist over time (*t*) with some factor (*p*) to damp out rapid fluctuations.

$$D_t = p*D_{t-1} + D_{in}$$

The robot should operate usually in the midrange of its capability as shown in Figure 8. Thus an optimal D should be MaxD/2.

If D is less than MaxD/2, then motor activation commands should be amplified by the difference. If D is greater than MaxD/2, motor inhibition commands could be amplified by the difference.

$$D_{out} = (1.0 - 2*D/MaxD)$$

D_{out} will range from +1.0 to –1.0.

D_{out} should affect all output elements equally. Although in the alternative implementation the influence may be indirect via the intervening excitatory (E) and inhibitory (I) elements (see Figure 8). For example:

If $D_{out} > 0$, then
$CL = D_{out} * \Sigma E_{CL} - \Sigma I_{CL}$
Else $CL = \Sigma E_{CL} + D_{out} * \Sigma I_{CL}$

As these output elements are coupled to mechanical process with inertia, the principal effect of changing D will be to change the overall rate at which the activity proceeds[12].

Various non-linear functions may be applied to smooth and limit this process. (see Blackburn, 1987).

4.9.2. Basic Reactive Pattern B. Balance

The objective of the BRP-B is to prevent falls and consequent damage to the vehicle. As the orientation and motion of the core will be the primary determinants of balance, balance may be assessed by the core accelerometers. The core will be taken along in many different ways, however, with the motions of the two track pods, but when these motions are expected, or predicted by the motor commands, they are at least purposeful, if not as dangerous as those occurring by accident. Balance or losing balance thus should depend upon whether the event was expected or not. To establish an expectation, the automatic control algorithms must make some predictions about how the core accelerometer data are going to change with a particular maneuver. If those predictions occur, then balance is maintained, however if events contradict those predicted changes, then balance would be upset. This prediction should be on-going and depend upon the integration of data from three vectors: 1) the pattern of activations of the different motors, 2) the pod leverage points, and the current conformation of the vehicle.

Balance would be calibrated continuously. This is most easily seen when the robot is planning to remain stationary on a stationary surface. Under that condition, its planning may involve nothing more than the absence of a decision to move. At this time the robot would be predicting no changes in its core accelerometers. Therefore, any change in the accelerometers indicates an unexpected change in balance, and should be met with a reactive and corrective response from its motors. Should part of the surface on which the robot is resting give way suddenly, the robot's accelerometer data would change as its core moves under the influence of gravity. This acceleration can be countered by activation of the track pod axes that would normally produce an acceleration in the direction opposite to the that of the fall, given its current configuration. The calculation needed is essentially an inverse of that used to predict a core acceleration in the particular

[12] The control processes should be less rate sensitive, and more position sensitive, so that an action will continue until completion before the next action is initiated.

direction of the error. For this calculation, we can use artificial neural networks that have a design rich in feedback. We will shortly describe a candidate neural network for this purpose.

While the expected pod accelerations for any activation of its motors is easily determined given fixed positions of the core, this will not ordinarily be the case with our NUGV. Nor will it ordinarily be our concern, as we need to predict the core accelerometer activity with activations of the pod motors. The core will be subject to disturbances caused by activation of any of the motors, and in all possible combinations while the pods themselves are at a great number of different positions with respect to the core and to their leverage points. Predicting the core accelerations is a complex multivariate problem.

The expected accelerations of the core will be functions of the motor commands of the vehicle (**M**), the present conformation of the vehicle (**C**), and the leverage points (**L**).

$$\mathbf{A}_e = f(\mathbf{M}, \mathbf{C}, \mathbf{L})$$

I have already defined the elements of the motor vector (**M**) and of the conformation vector (**C**).

The leverage point vector (**L**) is simply a feature set from the collection of data points from the sensors that detect and locate pod contact. The sensors that participate in this collection are the pod whiskers and the pod plate pressure sensors. For illustration purposes, let us assume that the contact profile for each pod was assessed by only four discrete sensors, each sensor either being on or off. One sensor would be located at each end of the pod, and one located on each pod plate. Then the pod contact could be determined to the resolution of those four locations by one of sixteen different features as shown in Figure 10. All conditions for each of the integrating elements *a-p* must be present before an output can occur. In the Figure, an input line terminating in arrow indicates the requirement for an active input, while the input line terminating in a dot indicates the requirement for an inactive input.

At this point, I should note that the sensor vectors undergo significant organization in most training algorithms for multi-layer perceptrons. An interim result of this organization is a vector of feature detectors similar to what I have shown in Figure 10. The network designer can greatly simplify the process of self-organization in a multi-layer perceptron by prescribing the connectivity that defines inclusively all of the potentially relevant features that are available from the sensor vector, even if some of those features are never used by the network in calculating the required (trained) output vectors.

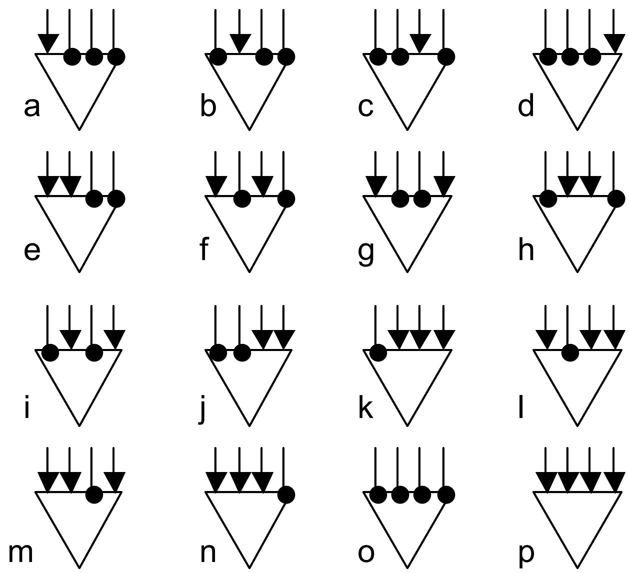

Figure 10. Firing conditions of sixteen hypothetical feature detectors.

The actual calculation of $\mathbf{A}_e$ can be performed rather quickly as long as the influences of the different motor commands given the different conformations and leverage points are known. We can discover these influences by observing what happens to the core acceleration under a variety of these conditions, and construct the gains of the function that calculates $\mathbf{A}_e$. This process is graphically demonstrated in Figure 11. The dotted circles in the Figure represent the conditioning signal. Facilitation is represented by line terminating in an arrowhead. Inhibition is represented by a line terminating in a dot.

Errors (**E**) in balance are detected by unexpected changes in the core accelerometers (***A***). The unexpected measure is a function of the difference between the expected (e) and the actual (a) reading.

$$\mathbf{E} = f(\mathbf{A}_e, \mathbf{A}_a)$$

In the process of defining the function that predicts core accelerations, the observed core acceleration with a particular controlled activation of the pod motors is compared with the current output of the integrator that develops the expected core acceleration. Initially there will be completely nonsensical prediction, and the error vector (E) will either look pretty much like $\mathbf{A}_a$ or like parts of it. This error is passed back over to the expectation integrator to modify the gains that determine the influence of its inputs. I give the modification rule in the section on *Activity Dependent Facilitation* later in this paper. Under controlled conditions, the network learns to predict what actually happens to the core accelerometers.

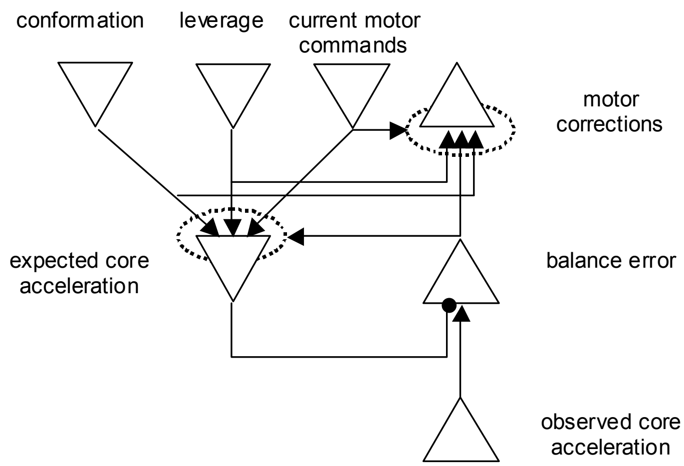

Figure 11. Method for the acquisition and execution of balance.

We are not finished with the balance calculations however, for the objective of the balance response is to restore the undisturbed position or orientation of the agent under uncontrolled conditions.

What the robot does to correct errors in balance should depend also upon its current state at the time the error occurred. Another way to look at the relationships in Figure 11 is to replace the data in the expectation vector with the data in the error vector and then ask for the motor commands that would contribute to that specific output vector. In other words, the detected error in acceleration could otherwise be a predicted core acceleration given a specific conformation of the robot, its current leverage points, and a pattern of motor commands. We need to solve a set of simultaneous equations to come up with the missing motor vector that would correct the balance error. We already have the necessary information to calculate the required motor correction to any error as we have calculated previously **C**, **L**, and **E**. But as we do not wish to aggravate the error by using the exact motor commands that would otherwise generate it, we simply need only invert the motor commands to activate an opposing response. This is easily done as our motor vector is composed of matched pairs of elements. We invert the vector by crossing the inputs of these matched pairs.

The factors involved in the calculation of the motor correction vector are also shown in Figure 11. Initially the error will be quite large and approximates the core accelerometer activity. Initially also the output of the motor correction integrator will be quite small as its input gains are undeveloped. Very quickly the system learns to predict core accelerations and to predict the motor commands that generate them. The system will thus learn to predict itself.

As the motor corrections are inverted, and the conformation and leverage inputs are continuous, this process in the absence of a core acceleration error, if left as described, could significantly interfere with on-going motor commands. The **C** and **L** vectors are necessary but insufficient conditions for a motor correction. There are various ways to inhibit the output of the motor correction vector until an error is present. The objective in all cases though would be to prevent the motor correction until the output of the error integrator represents primarily errors of prediction, and the output of the motor correction integrator represents primarily error specific responses.

In summary, the changes in motor commands necessary to correct for any disturbance in balance are a function of the delta in the balance vector that describes the nature of the disturbance and the current state of the robot that will determine how it can best respond to the disturbance. Both the prediction of the core acceleration with any given conditions, and the motor commands that would generate core accelerations with any given conditions can be acquired from experience under controlled circumstances. Balance errors would be corrected then by sending inverted motor commands to the twelve output elements. In early development of our control algorithms we can use the observed core accelerations under stable conditions to define (condition by experience) the transfer functions between the **C**, **L**, and **M** vectors and the $\mathbf{A}_e$ vector. As the network learns to accurately predict the core accelerations, the error between predicted and observed decreases and the output of the error integrator drops to zero, until an unexpected event occurs. Simultaneous with the self-organizing process of core accelerometer prediction, the motor corrections required to restore balance given any particular balance error (**E**) are conditioned by the current motor commands. As each motor is always subject to opposing commands, the inputs to the opposing motor integrators are conditioned by the current motor activity. Motor correction commands would normally appear after conditioning only when a balance error occurs. The inhibition of the balance error integrator by the expected core acceleration is very important to allow the execution of proper motor commands. I describe conditioning in greater detail later in the section on *Activity Dependent Facilitation*.

Quite often we should expect that the robot will lose balance when the unexpected event is caused by a state change in the external environment. Any attempt by the robot to restore its balance subsequent to this external environmental state change should fail. Fortunately, the robot has additional Basic Reactive Patterns that would be invoked in this circumstance, such as the BRP-D, and that could cooperate with the BRP-B to stay the unexpected accelerations.

The Japanese humanoid robot projects have developed algorithms to maintain balance in a multi degree of freedom robot. These methods should be studied for application here.

4.9.3. Basic Reactive Pattern C. Core Collision Avoidance

BRP-C prevents collisions of the core faceplates with objects in the external environment. The core faceplates contain video cameras, IR proximity sensors, whiskers, and SONAR. These sensors are open to the environment and, excepting the whiskers, must be protected. Activity from any of those four sensor types could be used to trigger the BRP-C, but initially the reaction will occur only in response to activations of the core faceplate pressure sensors. That is, the basic reaction will be a response to actual collisions, rather than a collision avoidance mechanism. We will introduce a method later in this paper that will progressively associate activity of the IR proximity sensors, of the SONAR, and of the digital video motion vectors with the more proximate detectors down to the whiskers or touch sensors. Only after the system has learned to associate events detected by the distance sensors with events occurring in response to activity in the sensors for the internal environment, true collision avoidance will be possible. The BRP-C backs the vehicle away from an actual (and later – impending) collision. The backing reaction can be a transient inversion of the preceding motor commands. To prevent the robot from getting stuck in an infinite loop, a random noise can be imposed on the subsequent forward command.

The BRP-C should help to reduce entanglements as the vehicle will avoid moving into objects that may get within the core-pod domain.

The Basic Reactive Pattern C is probably the simplest BRP to explain and to implement. Only the core face-plate sensors will trigger this BRP, and the response pattern will usually depend upon the on-going motor commands[13]. For example:

If W_f, then reverse current motor commands.

Where W_f is the core faceplate whisker(s), pressure, or touch sensors in the forward direction of travel. In lieu of whiskers, any strain sensitive device attached to the faceplate/core juncture could serve as the detector for collision.

The faceplate protection response offers an opportunity for the director to easily inhibit any ongoing activity of the robot. Lightly tapping on the faceplate should reverse the activity of the robot[14].

4.9.4. Basic Reactive Pattern D. Track Contact

The objective of the BRP-D is to optimize track contact with leveragable surfaces. From our earlier discussion of the Fixed Action Patterns, we should conclude that the robot prefers a conformation in relation to its leverageable surface in which the plate pressure

[13] The response is complicated a little bit when distance sensors predict a collision and where the robot may not be in motion, or may be moving too slowly to escape the collision. In this case, the distance sensors must be mapped to the motor commands that are associated with receding objects in that sensor field. This mapping can self-organize with the robot's experience. We will take up the possible mechanism later in this paper.

[14] During training (to be discussed in the section on Learning and Adaptation), tapping on the robot's core faceplate could punish a behavior as effectively as a collision.

sensors are maximally active. The BRP would press the track upon a surface that is perceived by the contact sensors (pressure plates and whiskers) to lie within reach. The infrared proximity sensors and SONAR may be conditioned to participate in this BRP based upon the response patterns controlled by the pod pressure plate and whisker sensors. Preferred surfaces would lie either below the pod track (in the direction of gravity) or in front of the track (in the direction of travel). Pod rotation and camber, and track tread rotation may be employed to achieve track contact. If no contact is made, the BRP-D should cause the pod to randomly explore its immediate environment in search of a contact point. If the whiskers indicate contact, then the BRP-D should move the pods in such a way that the whisker contact is replaced by a contact with the center of the tracks. When track contact is made, the BRP-D should attempt to enlarge it.

Thus, the BRP-D has three components. The first component addresses the means to make track contact, any contact. The second component addresses the means to shift the contact point from the location of the whiskers to the location of the center of the track. And the third component addresses the need to increase track contact.

Exercise of the first component should cause the robot to find and press upon the arms of a person who was unlucky enough to be suspending the robot by its core[15].

Exercise of the second component should facilitate the climbing of the robot upon any obstacle that it encountered or that was placed in contact with its whiskers[16].

Exercise of the third component should prevent the robot from rolling into an abyss, and should complete the effects of the first two components.

Similar to the determination of an appropriate motor command vector to restore balance, the BRP-D will use the leverage vector (**L**) to assess pod contact and the conformation vector (**C**) to assess the robot's conformation. However, it will assess **L** for each pod separately. Thus there will be a **LL** and a **LR** for leverage of the left pod and leverage of the right pod respectively. Recall that the **L** vector will be composed of features that individually describe the conditions that will determine which of the three components of the BRP-D should execute. It will further assess the activity of the individual elements in the **LL** and **LR** vectors. The three running questions for the BRP-D controller will be the following:

1. What modifications to the present motor commands will be necessary to randomly search the physical space for contacts?
2. What modifications to the present motor commands will be necessary to replace whisker contact points with track contact points?
3. What modifications to the present motor commands will be necessary to increase the magnitudes of the pod pressure elements of both **LL** and **LR**?

---

[15] The grasping reflex of primates is a biological example of this BRP-D component.
[16] A similar response will be observed when one strokes the breast of a perching bird.

The answers to these questions will involve first the assessment of the current **L** vector. We can do this by way of example using the data from Figure 10. Let us assume that the four pressure sensors come from a pod that is connected to the forward end of the core, and is situated in the closed position on the left side of the core. Let us also assume that the input pattern from left to right on each element of Figure 10 represents the four pod pressure sensors distributed clockwise on the pod starting from the top, then fore, bottom, and aft positions from the perspective of the core. Then the element *c* in Figure 10 represents the pod making contact only with the surface upon which the pod is resting. Element *o* represents a complete suspension of the pod with no contact points except for its connecting axle with the core. Element *f* indicates that the pod is wedged between contacts points on its top and bottom surfaces.

BRP-D Rule #1: If the active feature in the **L** vector for that pod is feature *o*, then the BRP-D would initiate a search of local physical space by activating the pod rotation and camber motors. The sequence and durations of the activation would be randomized[17] and could persist until a contact was made (a feature other than *o* occurred in **L**), or until a time-out was reached, or until balance was significantly disturbed.

BRP-D Rule #2: If a contact was made that resulted in the features *b* or *d*, the BRP-D could activate a pod rotation command and suspend any camber motion until features *a* or *c* appeared.

BRP-D Rule #3: If a contact was made that resulted in the features *a* or *c*, the BRP-D could continue the pod motor activation as long as the strength of the input was increasing. If the strength of the input began to decrease the pattern of pod motor activation could be reversed, and the test repeated. Both camber and rotation degrees of freedom would have to be tested separately. During running (FAP-R) the track rotation motors should also be subject to this rule.

The above rules will account for most of the conditions occurring during performance of the seven Fixed Action Patterns. Occasionally, the **L** vector will contain the features *e*, *g*, *h*, and *j*. These will represent encounters with obstacles while moving. Obstacles can be handled in two basic ways: 1) scale them, and 2) go around them.

The FAP-S and FAP-W will tend to favor the first option.

There remain several other conditions in which features *i*, *k l*, *m*, *n*, and *p* could occur. These conditions would involve, in general, an entrapment of the pods. To escape from such conditions, the robot may best run through all three rules randomly and vigorously, until freedom is achieved. To incorporate this strategy into our rules, we need only add

[17] Adding random noise to a control process in the absence of sensor input may not be necessary for, in general, sensor noise exists in all sensor systems, and is ordinarily suppressed by the presence of valid sensor input. Local adaptation mechanisms that adjust the sensitivity of the sensors are the usual means of this suppression. As the valid input decreases, the sensitivity goes up and the probability that noise reaches a stimulus threshold becomes one.

the conditions to the initiation list of Rule #1. If the motions of the vehicle generated sufficient wiggle room, then the execution of Rules #2 and #3 would be possible.

The activities of each pod are governed by these rules based on the local pod contact sensor information. However, some effects of any activity of any portion of the robot will be transmitted to other portions of the robot, affecting their dynamics, their leverage, and their sensor feedback. In a sense the two pods will compete for optimal contact with leverage points, transmitting their intentions through their physical connections with the core. The BRP-B will serve to mediate any conflicts.

The robot will usually have some physical reference (contact or leverage point) during translation. Therefore, if it encounters an abyss the BRP-D should prevent the robot from dropping into it. Instead of moving into an abyss, the BRP-D should reorient the robot to continue along the surface on which it has established leverage for its motion. This should happen because the BRP-D attempts to increase track contact. Movement that decreases track contact would be quickly interrupted.

Running along a beam or branch is a simple modification of FAP-R by the Basic Reactive Pattern D for Track Contact.

4.9.5. Basic Reactive Pattern E. Energy Level and Use

The objective of BRP-E is to acquire and conserve energy. The sensors for BRP-E measure energy reserve, and energy consumption or utilization. The homeostatic tolerance for energy level is quite broad, and describes a Sigmoid similar to that for Inhibition in Figure 9. Energy acquisition behaviors need be triggered only when energy reserves are quite low. In general, the detection of low battery charge should interrupt most on-going behavior, and trigger a recharge-specific behavior[18]. In the natural environment, with a limited or non-existent repertoire of navigation behaviors, the energy-limited robot may best stop all random motor activity and broadcast a call for help.

4.10. Motivation

The sensor inputs that govern each of the five basic reactive patterns above are analogous to biological motivators. So, for lack of a good engineering term, I call them *motivators*. Once again the five basic motivators are activity, balance, core collision avoidance, track contact, and energy level.

Table 2 reviews the relationship between the short list of behavioral constraints, which serve as the intrinsic motivators that drive and determine the most appropriate robot behavior, and the sensors that monitor the robot's internal state space (interoceptors).

[18] In many interior robotic systems, an example of a recharge-specific behavior is for the robot to home on its charger and plug itself in. This would be a little more difficult to accomplish for an exterior robot operating in a complex natural environment.

The Activity motivator is biphasic as it may either increase or decrease activity; the Balance motivator is monotonic and quickly affects activity to restore balance; the Core Collision Avoidance motivator is also monotonic and quickly affects activity to withdraw from collisions; the Track Contact motivator is biphasic as either the absence of contact or the extremes of contact generate a quick search for a preferred contact profile, while the occurrence of a preferred contact profile generates a slower attempt to optimize it; the Energy motivator is currently monotonic as low levels of energy reserve trigger only energy conserving activities. We may be able to show later that as the energy reserve moves closer to that trigger point, other energy acquisition behaviors might be invoked.

Motivator	Influence	Supporting Interoceptors	Utility to Robot
Activity (A)	biphasic	Core and Pod accelerometers and Pod track rotation sensors	Assesses the result of movement commands. Keeps the robot working.
Balance (B)	monotonic	Core accelerometers	Maintains orientation with respect to gravity. Assesses the result of movement commands.
Core Collision Avoidance (C)	monotonic	Core faceplate contact sensors	Protects distance sensors. Reduces friction Prevents entrapments.
Track Contact (D)	biphasic	Pod whiskers, Pod plate pressure sensors,	Localizes leverage points.
Energy Level and Use (E)	monotonic	Battery charge and current	Maintains adequate energy

Table 2. Basic Motivators

4.10.1. *Homeostasis Is Represented by Certain Values of the Motivators*

When the motivators are in the ranges proper for optimal performance, homeostasis is achieved. We may define a scalar variable for homeostasis (H). Then, the magnitude of H may represent the totality of the motivator states.

$$H = f\,(A, B, C, D, E)$$

We may arbitrarily specify at what magnitude of H is desired, then, in psychological terms, achieving that magnitude would indicate a comfort zone, and diverging from that magnitude would indicate disphoria. We will use this measure of homeostasis later in our discussion of learning and adaptation.

4.10.2. Motivation is always referenced to a particular process

A sensor value that is out of the "comfort" range for the particular process monitored by the sensor, will contribute to an increase in motivation to restore that particular process. Prior to the acquisition of significant experiences, we can reference any change in the behavior of the agent back to a particular interoceptor value that is out of its comfort range. The "out of comfort" sensor may also contribute to the more general disphoria measure *H*, depending upon the overall level of *H* achieved.

4.10.3. Deprivation and Excess Determine the Strength of a Motivation

The magnitude of the deviation determines the degree and direction of the motivation. Following are some examples.

> If track contact was absent, the drive for track contact would rise and contribute to increasingly greater motions to establish a contact and restore leverage (and simultaneously a stronger inhibition of the track rotation until adequate leverage was re-established).
>
> If the robot was restrained, the drive for movement mediated by BRP-A would increase and the motor responses with all other BRP would be amplified.
>
> If the surface on which the robot was resting pitched to and fro like the deck of a ship in rough seas, balance would be repeatedly challenged and the motivation to stabilize itself would increase. The response to repeated loss of balance should include, in addition to an attempt to increase track contact with leverageable surfaces, an attempt to close upon any available leverageable surface. This should occur as a consequence of the cooperation between BRP-B and BRP-D.
>
> If the core faceplate sensors were repeatedly activated, the ongoing motor behavior could be more vigorously reversed. This should occur as a consequence of the cooperation between BRP-C and BRP-A.

4.10.4. Governing Sensors Cooperate to Invoke Appropriate Behaviors

The examples above also illustrate how motivators cooperate to invoke and control the most appropriate behaviors. Here are a few more examples specific to this point.

> If the tracks reported contact and were spinning with no apparent accelerometer indications of forward motion of the robot core, the activity motivation would increase. This would trigger an action pattern transition from FAP-R to FAP-W.

If the robot was tumbling down an embankment, the drive for activity would decrease, yet balance would be disturbed and the robot would likely invert its motor commands in an attempt to counter the motion.

Depending upon the axis on which the robot was tumbling, and the concomitant sensor activity, the robot could transition to different conformations. If the rotation of tumble was on the X-axis (see Figure 6), the robot could initiate a FAP-S due to its obstacle scaling response. This would open the pods from the core and oppose the tumble.

If the robot was tumbling on the Z-axis, the robot could initiate a turn response to avoid obstacles detected in its pod proximity sensors. This response should also oppose the energy of the tumble.

4.10.5. Self Awareness

The robot will be self aware to the degree that it can optimize its homeostasis. Awareness, like perception, requires not only sensor processing but also an effective motor response. On the sensor side, the sum of the information from the interoceptors of Table 2 constitute the input for self-awareness of the robot. We have shown how the robot must assess and integrate information from all accelerometers to make a determination of its current conformation and of how its conformation is changing. In addition, the robot uses the accelerometer data plus the track velocity sensors to assess its motion with respect to its leverage points. The robot will use the H measure above as a general index of comfort or of its inverse – disphoria, while the specific sensors monitoring the critical state variables will provide information on what must be addressed at any moment in time. On the effector side, the robot will use its six degrees of motion freedom to avoid disruptions to homeostasis and to restore the critical state variables.

4.11. Beyond the Fixed Action Patterns

We should ask at this point in our discussion of just of what is the robot capable? Given only the five Basic Reactive Patterns and the seven Fixed Action Patterns we expect that the robot could self-initiate activity as its motivation for activity would initially be quite strong. We should also expect from Table 1 that the first FAP to be assumed would be the Run. Other FAPs may follow as conditions warrant. But Run to where? An agent with very poor external sensor capabilities may best move randomly through the environment, and depend on its Basic Reactive Patterns to keep it out of trouble. Eventually though, our robot would run out of energy. The high probability for this catastrophic event is due to our design omission that does not provide the robot an opportunity to acquire energy during any FAP.

4.11.1. Motivation is Necessary but Insufficient for Reliable Survival

By responding appropriately to the five basic motivators, the agent may survive transient challenges to its homeostasis brought on by the execution of any of its seven Fixed Action Patterns, but it would yet tend to be subject entirely to the fluctuations in its external environment. One mechanism that nature has successfully employed to reduce this environmental subjugation, is to employ distance sensors and associate subtle changes in the external environment with significant consequential changes in the internal environment. Upon detection of those subtle changes in environmental cues the agent can invoke a reactive process that either avoids or approaches the environmental cue. Those cues that are associated with events that restore or maintain homeostasis are fortunate for the agent. Those cues associated with events that do not must be avoided, otherwise those events will tend to terminate or exterminate the agent. Therefore, we must provide sensors of the external environment that will detect with sufficient sensitivity the subtle changes (the cues) that will predict significant change to the robot's internal environment, and we must provide a mechanism by which the robot can determine the most appropriate way to respond to those external events.

4.11.2. Sensors of the External Environment

Certain sensors that monitor conditions in external environments are installed on the vehicle. These are four IR short-range proximity sensors, four mid-range SONAR, four color video cameras, four acoustic microphones, and two RF transceivers[19]. Table 3 lists the external sensors and possible low-level uses of the available information.

Sensor	quantity	locations	Range/sensitivity	Applications
SONAR	4	Core face plates/pods	12 < r < 48 inches	Distance to obstacles/leverage points
IR	4	Core face plates/pods	< 12 inches	Distance to obstacles/leverage points; presence of warm objects
Video	4	Core face plates		Color of objects; object distances from optic flows
Stereo Audio	4	Core side plates	> X dB	Relative location of activity
RF	2	Core top plates	1000 feet	Direction to OCU; communication with the director

Table 3. Sensors of the External Environment

The basic purpose of these external sensors is prediction. To improve upon its homeostatic mechanisms, the robot may use its external sensors to predict the different conditions that it will encounter during its movements. We have noted that the robot's

[19] The sensor side of the RF transceiver is the receiver that accepts (senses) communications from the operator control unit.

movement through the external environment engenders certain risks. Such risks are primarily related to collisions and to loss of contact with leverageable surfaces (e.g. falls). The external sensor information then should presage those hazards. Also, the movement of the robot may increase its likelihood of being recharged. The external sensors should detect the critical environment features that are associated with an energy source[20]. Similarly, movement itself is a homeostatic motivator, thus the external sensors should provide information that will indicate a traversable pathway (that is, one that does not impede movement).

The robot has little control over its external environment, yet its movement within that environment can change the impact that the environment might have upon it. For example, the external sensors might detect a looming object and the robot could predict a possible collision. The robot could move out of the way using similar behavioral strategies to those that it would employ had the collision been a result of its own motion through a static environment. Its avoidance of the looming object might preserve its own physical integrity, but have no effect upon the trajectory of the looming object.

Earlier in our discussion of the Fixed Action Patterns I indicated how the different patterns could be invoked by activity in the interoceptors. Ideally, the exteroceptors will provide predictive information that can be used to invoke the transformations among the Fixed Action Patterns in advance of the interoceptor triggers. In both cases, the changes in behavioral patterns should be appropriate for the conditions in the external environment, but in the second case, the robot could anticipate changes in the external environment and prepare for them. This could reduce errors and increase the speed of activity.

In a competent local control process, the Fixed Action Patterns should vary with environmental conditions, detected by the external sensors, and modified both by additional external sensor information and by the internal sensor information that continually attempts to optimize homeostasis. For example, the robot may detect an approaching object through its stereo audio and video sensors. The robot could move away from the object provided that it had adequate energy reserves and a navigable path to follow. If the range sensors indicated that the path ahead of the movement was clear, the robot may initiate and continue in a FAP-R. If the SONAR indicated an obstacle ahead, the robot could turn in the direction of the clearest path as indicated by its side looking SONAR and the optic flow from the peripheral fields of its forward-looking video. If the rearward sensors indicated a progressive pursuit, and the forward sensors indicated a proximate obstacle, the robot could shift to FAP-W, and attempt to walk over the obstacle. If the obstacle caused a total tilt greater than approximately 30 degrees, the robot could initiate a FAP-S, but if the total tilt angle was greater than approximately 45 degrees the robot could suspend pod rotation in the extended position and then continue to ascend or descend in the frozen FAP-U pattern. However, if the obstacle turned out to be steep but short, the robot could suspend FAP-U when the core was maximally

[20] For example, as the human operator is most likely to be associated with energy recovery, the robot could associate the features indicative of the presence and location of a human operator with its energy motivation, and orient to those features when energy reserves were low.

elevated, and then shift to FAP-R to move without further undulations but with a higher perspective over the terrain. Tilt of the video cameras could be accomplished by differentially rotating the pods in FAP-U.

5. *Learning and Adaptation*

5.1. Motivators Protect, Prioritize, and Reinforce Behavior

The motivators of the five Basic Reactive Patterns can serve many functions in the control of behavior. Not only do motivators trigger and govern reactive behaviors that provide immediate protection for the agent, they can also serve as mediators to determine which of many competing behaviors are selected for expression, and they can serve as the criteria for the acquisition of new behaviors. We have already seen examples of the first two roles, I will next address the mechanisms of learning and of decision-making, and explain the third role for our motivators in the control of behavior.

5.2. Learning Enables Prediction

It is axiomatic that the measure of success for learning (long-term adaptation) is the restoration or maintenance of homeostasis. Learned behaviors are appropriate when they promote the welfare or survival of the agent, which are possible only under homeostasis. For our agent, the Novel UGV, survival may be determined by the availability of energy, by the continued operation of its hardware and software, and by its utility to the human operators. When utility disappears, the agent is subject to the trash heap. When energy dissipates, or when functionalities of hardware or of software cease, the same trash heap awaits. The learning objectives then, from the perspective of the agent, should be to maintain its energy reserves, keep itself together and functional, and meet the needs of its user. The reader may note that this last objective is something new compared to the five basic motivators discussed earlier. What will make this new objective possible is learning and long-term memory.

The external sensors provide information on the environment that can be used both to predict a homeostatic catastrophe and to predict behavioral alternatives that, if taken, will avoid catastrophe. Learning is the device used by adaptive natural agents to predict the conditions in the external environment that will have an impact on the internal environment and change homeostasis. By reacting to the predictions of these environmental conditions, in advance of their impact on the internal environment, an agent is more likely to maintain its homeostasis. We will emulate this device in our artificial agent, the Novel UGV, to provide it with a similar advantage.

We now describe the learning mechanism that will associate the information available from sensors of both the external and internal environments, predicting their homeostatic consequences, and directing future behavior to avoid or approach those external factors given the current internal state.

5.3. Classical Conditioning

During performance of a Fixed Action Pattern, the Basic Reactive Patterns will modulate the motor commands according to rules implemented in the Fixed Connection Matrix.

These rules are analogous to the unconditioned stimulus-unconditioned response pairings of classical or Pavlovian conditioning. When the robot is able to perceive features of the external environment through its distance sensors, this information becomes available for association with the unconditioned response. During movement, the core accelerometers, pod pressure sensors, and faceplate pressure sensors provide the major unconditioned stimuli to support (facilitate) a conditioned response of features from the distance sensors. After conditioning (the repeated co-occurrence of the internal and external events), the features from the distance sensors invoke a response similar to the unconditioned response but in absence of the event that originally produced it. The classical conditioning paradigm is diagrammed in Figure 12. In the sequence of events during conditioning, the external event usually precedes the internal event (a likely happening because the external sensor is a distance sensor), but the record of the occurrence of the external event persists if not the event itself. When the internal event occurs it evokes a predictable response to restore homeostasis. The persistent trace of the external event becomes associated with the response evoked by the internal event according to the mechanism of *activity dependent facilitation*.

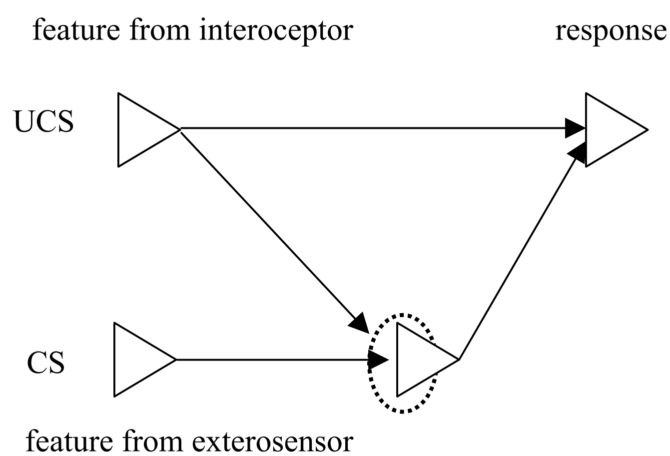

Figure 12. Simplified Classical Conditioning Paradigm

5.4. Activity Dependent Facilitation

A general learning law, known as activity dependent facilitation (Kandel and Hawkins, 1992), approximates classical conditioning and is useful in determining the contributions of a particular input through its modifiable connection to an integrating element preceding an output decision. The law is as follows:

$$\Delta w = G * ((z/e) * w) * (a *(S - m) *(C - w) - m *(w - c))$$

where w is its current connection strength, z is the activity on the input element in question, e is the sum of inputs from all cooperating elements to the integrating element (prior to their filter by the *w* vector), *a* is the total activation of the integrating element (equivalent to the product of the *e* vector and the *w* vector), *S* is a constant representing the maximum permissible sum of weights connecting to any one element, *m* is the current sum of weights making contact with the integrating element, *C* is a constant representing the maximum permissible weight, *c* is a lower limit on the weight to prevent it from disappearing completely if rarely used, *G* is a constant = 1 / (S*C). When both *z* and *a* are present, *w* is increased, but when *z* appears alone, *w* is decreased.

The influence of the unconditioned stimulus in the above learning law is incorporated into the sum of inputs on the integrating element. The connection weights for the UCS are strong, not modifiable over the short term, and reliably invoke an output decision in the absence of any other cooperating inputs.

The use of this law permits associations among previously ineffectual feature vectors, so that several layers of conditioning can occur. A general method of classical conditioning using the above learning law and that provides for the evolution of behavioral sequences is schematized in Figure 13. The method includes an input field of the observed sensor patterns, short-term storage of the history of those patterns (analogous to short-term memory), an output field of the predicted pattern that should accompany the result from the associated motor command, a comparator of the expected and observed patterns, a field to temporarily store the resulting errors, and an association matrix of the input history with the current input, the current error and the next motor command. Feedback is completed through the external environment. In Figure 13, only two processing elements with all of their connection are shown in each field for clarity. The actual numbers of processing elements in the different fields depends upon the resolution of the sensory field, the complexity of the effector (motor) system, and the resolution and complexity of the feature detectors. During conditioning, the UCS for matrix A is a collateral from the Base Reactive Pattern that is currently in effect. The UCS for matrix B is the current sensory input, and the UCS for matrix C is the current error. In each case, the UCS is the event to be predicted. Using an algorithm similar to the model in Figure 13, a predicted motor response will execute in advance of the original BRP.

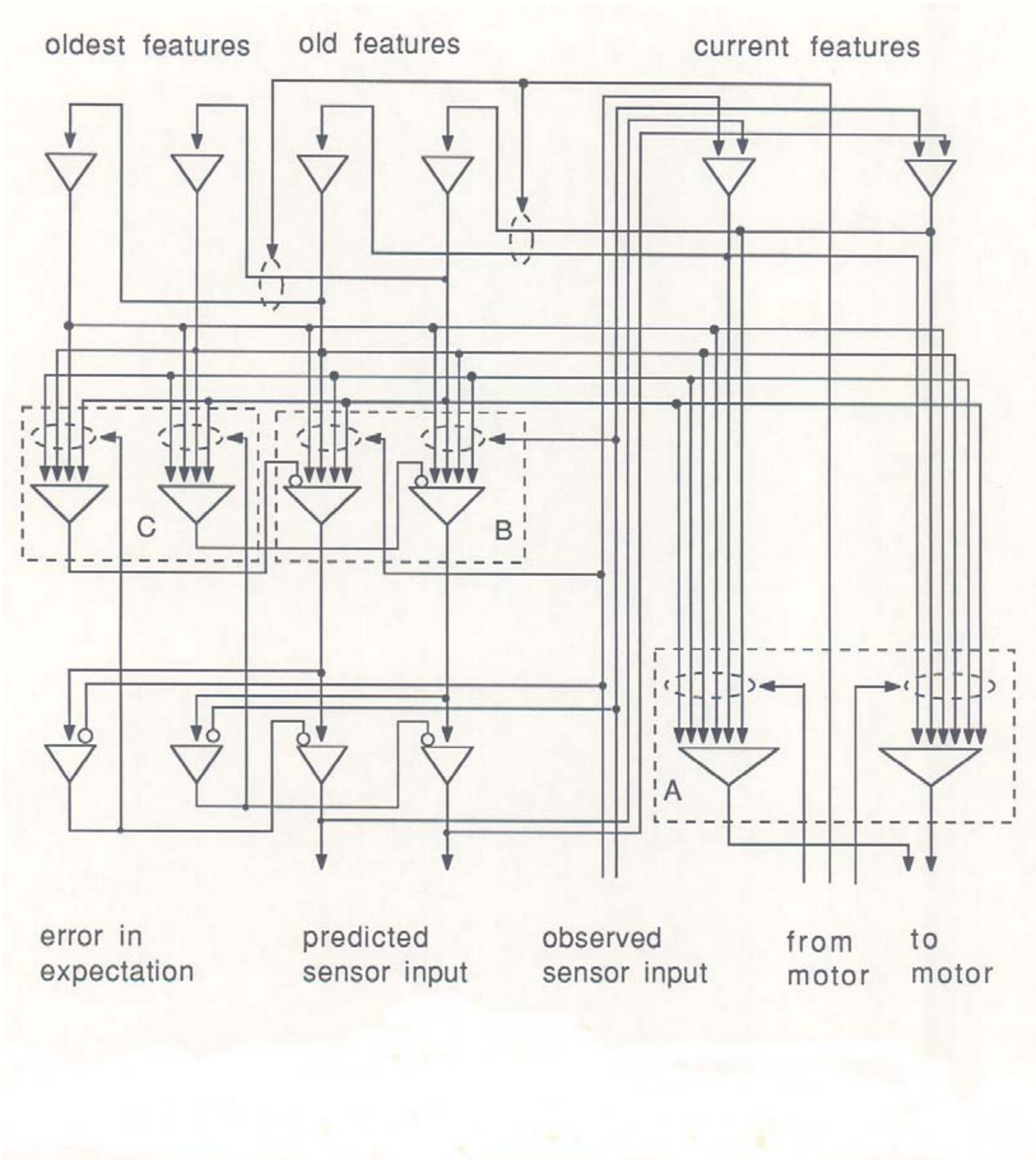

Figure 13. General Model for Classical Conditioning of Perceptual Motor Sequences

In classical conditioning, novel information from the external environment acquires the strength to evoke responses that already exist in the agent's repertoire and are appropriate for the general conditions that the novel information predicts. Additional information on the application of this learning model is available in Blackburn and Nguyen (1994).

5.5. Operant Conditioning

The post-hoc appropriateness of any particular behavior is determined by factors that change the sensor values, and, in effect, indicate the change in probability of catastrophe. Our second axiom is that the Basic Reactive Patterns of behavior operate to reduce the probability of catastrophe. Thus, the Basic Reactive Patterns show the Adaptive Behavioral Controller how to operate in order to restore homeostasis. That is, when a behavioral action initiated by some command from the Adaptive Behavioral Controller results in an internal sensor reading that indicates that a) activity is restored to its midrange, b) balance is restored, c) collisions are avoided, d) track contact is improved, and/or e) energy reserves and/or energy conservation are improved, we can be assured that the probability of a catastrophe has been reduced. These successful behaviors under the given environmental conditions, should be remembered so that they can be repeated whenever the appropriate conditions reappear. Similarly, when a behavioral action results in too much or too little activity, loss of balance, collisions, loss of track contact, and depleted energy, that behavior should also be remembered **and inhibited** whenever those prevailing environmental conditions reappear[21].

When a specific motivator is out of homeostatic bounds, the previously associated behaviors should be primed for action. An efficient way to accomplish this priming is through the association of the interceptor features with features from the exteroceptors. The biasing of the exteroceptor features would in turn bias specific behaviors when the environment contained stimuli characterized by those features.

Our third axiom is that all acquired behavior for our robot will be expressed through the modulation of the seven Fixed Action Patterns using pathways in parallel with the five Basic Response Patterns that also modulate the FAP.

Recall that the FAP are generally modifications of the FAP-R that is executed while the robot is in its normal closed conformation. The robot expands from this conformation to adapt primarily to information from its immediate neighborhood sensed by the IR and Whisker sensors. Recall also that the BRP generally motivate and modify the FAP based upon information from the sensors that are monitoring the internal environment. Thus, through our external influences on the stimuli that control the BRP, we can intervene and modulate any motor command associated with any FAP during performance.

Evidence that learning has occurred will be a modification of a FAP that is not immediately predicted by a complete knowledge of the internal and external environment, for learning will have permitted the robot to predict and precede an environmental event with a unique behavior.

For those readers familiar with the biological Learning Literature, we will implement here analogues of instrumental (or operant) conditioning, also known as *reinforcement learning*. Like classical conditioning, reinforcement learning requires the agent's perception of environmental information. In addition, operant learning requires an action on the part of the agent separate from the unconditioned response, and it requires some perceivable consequences of that action. The agent can use any of the available sensor information for the assessment of the environment and for the assessment of its behavioral consequences.

[21] The exclusive-or problem that is solvable by a three-layer perceptron is an example of a two choice paradigm where one choice must be inhibited in favor of the alternative under the co-occurrence of two otherwise permissible stimuli.

The process and rules of reinforcement learning that we can implement are as follows:

- Assess the internal environment (I)
- Assess the external environment (E)
- Perform an action (A)
- Reassess the internal environment, and determine if homeostasis (H) is improved.
- If H is improved, then associate factors I, E, and A, such that if I and E, then facilitate A.
- If H is worsened, then associate factors I, E, and A, such that if I and E, then inhibit A.

The above rule suggests that our controller have a special circuit that can inhibit or veto a particular action. This circuit may participate in the association rule above whenever homeostasis is disturbed by a behavior. The rules for operant conditioning are graphically represented in Figure 14. In Figure 14, arrowheads indicate direction of information flow. The line terminating in a dot represents inhibition. The dotted circles represent locations of activity dependent facilitation or inhibition.

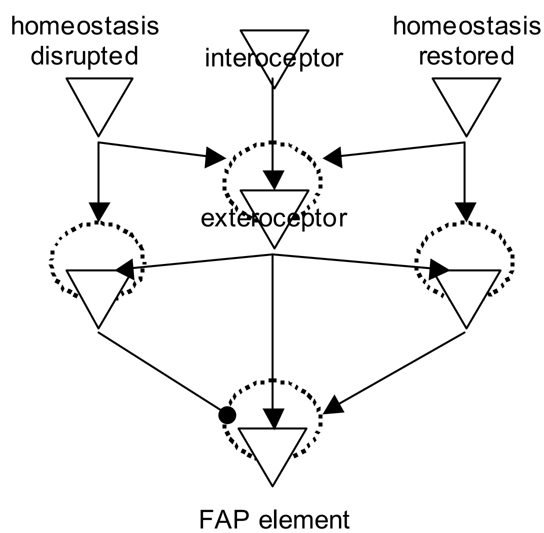

Figure 14. Simplified Operant Conditioning Paradigm.

The director, serving in this case as the supervisor of learning, need not go to great lengths to manipulate the environment in order that specific changes in homeostasis accompany particular actions under those conditions. This is because the learning algorithm above guarantees that the probability of occurrence of a particular action in the future will depend upon the prevalence of those specific internal as well as external environmental conditions. In the future, when the director may wish to see that particular action in response to particular external conditions, the internal conditions may not be present with sufficient intensity to drive the action above behavioral thresholds or above competing behaviors. Thus the director should generally not mess with the internal conditions of the robot.

The locus of learning in our control architecture of Figure 5 is the box labeled *plastic connections*. The reader will notice that this box receives input from the internal sensors, the external sensors, and the box containing fixed connections. The internal architecture and processes of the two boxes containing fixed and plastic connections respectively are not yet fully explained. We will define the connectivity within these boxes based upon the principles contained herein and report on these details in subsequent documents.

In operant conditioning, novel information from the external environment acquires the strength to evoke responses that already exist in the agent's repertoire, but that were previously unrelated to any intrinsic motivators.

5.6. Fixed Action Patterns Provide the Basis for New Behaviors

The *reinforcement learning* algorithm above requires that an action take place before the test of homeostasis. Before learning, the only behaviors of which the robot is capable are the fixed action patterns. Thus the robot will be performing a fixed action pattern when learning initially takes place. Learning will modify the particular FAP and invoke that modified FAP pattern in the future whenever the associated internal and external environmental conditions are present. When the environment is novel, the agent will default to previously learned behaviors or to the original FAP, depending upon the degree of novelty and motivation.

After some modifications of the seven FAP, the repertoire may be expanded with newly acquired behaviors by building upon the previous action patterns that are invoked by the prevailing environmental conditions. This process is known as behavioral shaping and permits learning to progress without destroying previously learned patterns. In this way the repertoire could become quite complex, depending upon the agent's ability to discriminate the necessary behavior specific features from the external environment, and upon its ability to respond differentially to those features.

The seven FAP exercise all of the mobility degrees of freedom of the robot in coordinated patterns that accomplish mobility under a variety of external conditions. The BRP provide transitory modifications to the coordinated FAP to meet certain exigencies and promote homeostasis. The external sensors can extend through classical conditioning the range of events through which the BRP are active. With any given external environment, and sufficient range of sensitivity in the external sensors, the modifications to the FAP, and even the switching among them, can create the impression of the invention of novel behavioral patterns, when in fact, only old patterns are being rearranged.

Classical and operant conditioning provide for one additional element that increases the potential for behavioral complexity and unpredictability give the immediate or current environmental conditions. That element is memory. Memory, however, is nothing more than the persistence of the associations between the features of the external environment and the features of the internal environment established through classical conditioning of the BRP, and through operant conditioning of the FAP.

5.7. The Selection of One Behavior from a Repertoire of Acquired Behaviors

Following learning, the occurrence of any desired behavior will depend upon not only sensor readings in both the internal and external environments, but also upon the configuration of the plastic connections, their current states, and their transition probabilities. The configurations and transition probabilities will depend upon the learning experiences, and the current states will depend upon the short-term history of activity.

Learning will release the agent from a strict adherence to environmental conditions. The motivators will continue to modulate behavior, and provide the fundamental drive, but the direction of the behavior will depend also upon previous experience. The confluence of the state of motivation and previous experience will add a degree of unpredictability to the controller relative to the information available to an external observer with knowledge of only the recent history.

Therefore, the agent will be able to select from a variety of potentially useful behaviors; the degree of utility will depend upon experience and the present conditions. The propensity to select from that repertoire will also depend upon experience and the present conditions. The rules that govern the selection and maintenance of a fixed action pattern are in fact the same rules that participate in the selection of a behavior from the available repertoire. Ideally, the robot would make no particular selection unless the conditions warranted it, but errors acquired in experience due to inappropriate reinforcement would surely result in errors in later performance.

When external conditions are insufficient to contribute to goal-directed behavior, a default behavior would likely emerge, for example, random search[22].

5.8. Energy as a Motivator and Shaper of Behavior

The presence of several cooperating behavioral criteria (see the list in Table 2) permit one or more criteria to emerge as the dominant driver of behavior and determinant of learning depending upon the conditions in both the internal and external environments. When energy reserves are high, for example, the detectors for movement may emerge as the dominant driver and not only select drive-specific behavior but also determine which novel behavioral patterns are facilitated and which are inhibited. When energy reserves are low, the change in energy reserve could be used to reinforce behaviors that contribute to energy acquisition, even if those might violate the criteria for movement.

[22] Random search could be appropriate when the agent is still acquiring a useful repertoire of behaviors. Afterwards, the agent may best meet the absence of requirements (i.e. operator commands, or energy disparities) with quiescence (but also with action readiness).

The robot monitors its energy reserves and attempts to maintain the reserves at an adequate level for continuous operation.

The robot can use energy reserves and energy consumption to control behavior and the acquisition of new behaviors. The replenishment of energy should be a strong facilitator of behaviors that led up to the event of replenishment. On the other side of learning, a high usage of energy during certain behaviors provides a cost measure for those behaviors that can be used to learn the avoidance of those behaviors in the future.

With respect to energy acquisition behaviors, a full battery capacity provides no particular stimulus or motivation for the robot, so the robot could be released from its focus on energy to do other things. A low battery charge, however, should trigger and maintain a set of behaviors that have proven through experience to increase the battery charge. Short of plugging itself into a power source, which may be quite distant and unpredictable, the robot must first acquire other environmental features that are most often associated with the acquisition of energy. In the biological learning literature, these other environmental features are known as secondary reinforcers. In our operational environment, the robot's director will most likely re-supply the robot with energy. Therefore, from the perspective of the robot, its director could take on the properties of secondary reinforcement. The director is like a *mother* to the robot, and we may look upon that relationship in very similar ways. Rather than seeking out new batteries directly, or wall sockets to plug into, or even a charging station, the robot may seek out its director with the expectation (implied) that the director will do whatever is necessary to recharge the robot's batteries.

An energy-depleted robot is probably useless for most of our applications. Thus we should arrange for the robot to seek out the human director whenever its energy reserves dip below some threshold. The threshold should be high enough to ensure that the robot can get back to the director, or at least to assist the director in recovering the robot[23].

5.8.1. The Robot Must Attend to Its Director

Next, we must address the question of how the robot will sense the presence or identity of its human director. This can be done in several ways, but each comes with some computational cost. Humans are unique, but the distinguishing features can be subtle. Face recognition and voice recognition may be useful, and technologically feasible. But at first we may be satisfied with only the robot's ability to detect where any human is located and to move in the appropriate direction to make physical contact.

5.8.2. The OCU as a Homing Beacon

All humans emit IR radiation, and usually have predictable body orientations and proportions. To use these features as cues, the robot must yet have some image capture

[23] One mechanism of adaptation is to vary the threshold for some decision. The threshold may be varied by the addition or subtraction of a quantity that is temporally and spatially consistent with the threshold in the decision process.

and analysis capability. But this too is beyond the current resources of the Novel UGV. Instead we might use the strength (and direction) of the radio transmission between the OCU and the robot as the director-defining and locating information for the robot. In other words, the active radio communication between OCU and robot can tell the robot that a director is accessible. To get direction from the OCU radio signal, the robot may need a directional antenna. The robot may change its orientation with respect to the RF signal until a maximum is found, then take that identified heading. For positive identity, the OCU might send an encrypted password that uniquely identifies the director.

5.8.3. A Process to Promote the Director as a Secondary Reinforcer

Initially the robot is indifferent to its director. However, because of the basic reactive pattern E, the robot is not indifferent to its energy reserve. When the energy supply is low, homeostasis is disturbed and the director has an opportunity to cause the reinforcement learning algorithm to associate that low energy reserve with environmental information that is unique to the presence and location of the director, and to an action that would bring the robot around on future occurrences of low energy reserves. The following learning paradigm might be employed:

- Begin training with depleted energy reserves
- Use the OCU on low broadcast power
- Approach the robot
- Apply external current to recharge the robot's batteries.
- Turn off the OCU radio.
- Turn off the external charging current.
- Repeat the process.

Following the above learning protocol that should install the director as a powerful secondary reinforcer for the energy motivator, the robot should have a propensity to seek out the director whenever its energy reserves are low. In this scenario, the director is synonymous with the communications signal. As processing power on the robot is improved, video and audio information can also be used to identify and localize the director.

The director can use his/her position as a secondary reinforcer to control additional learning. First the director would start the robot in a reduced energy state. This would trigger the director seeking behavior. Next the director could place obstacles in the path of the robot, over which the robot must learn to traverse. The increases in signal strength (from the communications signal, the audio signal, or the IR) could be used to intrinsically reinforce the behaviors employed by the robot in its traversal[24].

5.9. Behavior is Multiply Determined

---

[24] Simultaneously, the co-activation of the BRP will condition the acquisition of obstacle negotiation behaviors motivated by the reinforcement from the communication signal strength (see the section on Learning Mobility).

Because energy and communication are critical to the utility of the robot, the propensity of the robot to roam, explore, return to the director, or perform some other routine to improve communications may be dependent upon the operational conditions of communication and energy[25]. Some of the possibilities are given in Table 4 below.

From the Table of Conditional Robot Behaviors, we should conclude that the robot will not likely return to the director until it has run quite low on energy. If it roams until it has expended approximately 1/3 of it energy reserves, then we should expect that it would take another 1/3 to get back. If we could induce the robot to explore or perform some other task objective before it has expended 1/3 of its energy reserves, then the energy threshold for the switch between exploration and return could be temporarily reset accordingly, permitting a longer task duration.

		Communications Signal		
		strong	*moderate*	*weak*
Energy Reserves	*Strong (3/3- 2/3)*	Deploy, roam, or perform task	Explore, or perform task	Climb to restore communications
	Moderate (2/3-1/3)	Explore, or perform task	Explore, or perform task	Climb to restore comms
	Weak (1/3-0)	Return on comms gradient	Return on comms gradient	Return on comms gradient

Table 4. Example of Conditional Robot Behavior

As noted above, the director can become an attractor for the robot. Thus the robot should have a strong propensity to seek out its director when its energy reserves are low. The communications gradient tells the robot where its director is located in general. Once back into the environment of the director, the directional information from the communications signal strength may then be supplemented with video or IR information to localize the director. Director voice signals detected from the robot's stereophonic microphones may also be used to localize the director.

5.10. Learning Mobility

Using the organic sensors for acceleration and orientation with respect to gravity, touch, track pressure, and magnetometry, the robot should be able to detect its movement, its conformation, and its orientation with respect to the earth's magnetic field, and with respect to objects against which it is leveraged.

The robot is then presented with an objective. I have described how one such objective can arise. That is the orientation and directed movement toward the robot's human

[25] Climbing could be appropriate when communications were lost, and when an elevation was detected to climb.

director when energy reserves are low. The methods of movement across an undefined terrain are left to the robot to determine. We should expect that a number of fixed action patterns can be selected in response to simple events encountered during transit. These alone could execute some routes if a path was not too difficult and sufficient time was allotted. Our intent in providing opportunities for learning and adaptation utilizing distance sensors is to increase the probability that the robot will negotiate more difficult paths and select more appropriate routes to traverse the intervening distances in shorter times than would be possible by a more random method.

5.11. Learning Decision Making

As the purpose of the robot is to move in a controlled manner through its environment, it must maintain a friction-supported contact with some leverage points. The pod plate pressure sensors provide the contact information, but the presence of friction must be assessed by other means. One method is to compare track velocity with accelerometer input under the conditions of applied force from the track motors. Table 5 gives some of the possible outcomes.

Applied Force on Track		*high*		*low*	
Track Velocity		*high*	*low*	*high*	*low*
Pod Accelerometer Output	*high*	adequate track friction	unlikely condition	inadequate track friction for pod momentum	inadequate track friction
	low	inadequate track friction for pod momentum	adequate track friction for pod momentum	inadequate track friction for pod momentum	adequate track friction

Table 5. Possible Interpretations of Applied Force vs. Observed Motion

The robot should endeavor to keep friction adequate for the current combination of the load and the applied force. Depending upon the present conformation of the vehicle, the robot could modify its conformation to improve friction. A change in friction would trigger a learning algorithm associating the previous behavior (conformation) with the perceived environmental conditions and the behavior that resulted in the new conformation (based upon the law of effect). The reader may recall some of the possible conformations from Figure 4.

6. Operational Implications

6.1. The robot must always be "on".

The first requirement for an independent agent is that it must remain always in the "on" state. If an on/off switch is provided, it should be left in the "on" position, and only switched to the "off" position when major electrical modifications are required. Techniques for "hot swapping" of batteries, motors, and electronic components, that selectively turn off only portions of the robot during repairs and maintenance, may make it possible to keep the total robot system always in the "on" state.

It may seem counter-productive to have a robot that is always "on". However, this does not mean that the robot must always be "on the move". Rather, the robot must be always ready to move, and may even move often if the conditions are appropriate – for example the robot might be located in a high traffic area and simply be "in the way".

Let us consider some of the other advantages, even the necessity, of a robot that is perpetually "on". One advantage the owner of the robot would gain by permitting the robot to remain "on" is that the robot could be continuously prepared for work. Another advantage is that an "on" robot could spontaneously become active and exercise its skills. This exercise could improve its adaptation to its work environment without the need for operator supervision[26]. An adapted robot may then be able to work independently of human control.

Operationally, we expect that the robots will be often in the company of humans. Humans move about frequently. They do not like to lug their equipment with them when they move. That is why the Army is asking for a very lightweight mobile robot. But regardless of weight, it would reduce human workload considerably if the robots could orient to their human operators and keep up with them when they did move. To accomplish this, the robot must be able to operate to the limits of the human operator's mobility envelope. This mobility envelope includes some of the following[27]:

- Rapid waking and activation
- Traversals on a planar surface at rates less than four minutes per statute mile.
- Traversals for one hundred feet across a horizontal four-inch beam.
- Vertical jumps over a seven-foot bar.

---

[26] Two disadvantages of a periodically active robot could be that the robot might wear out its mechanical apparatus, and consume more energy than one that remained mostly in the "off" condition.

[27] Perhaps the best way to get a good impression of the limits of the human mobility envelope is to watch the series of events at world-class track & field and gymnastics competitions. As an alternative, and one with unequivocal military significance, is to examine the mobility required by a basic training confidence or obstacle course. These courses stress the strength and agility of young recruits without the benefit of levers or cushions. The course requirements are established not only to test the physical fitness of the recruits, but also to assess the readiness of the recruits to meet actual operational conditions. If we intend for our robots to accompany the operators in the field, the robots too should meet those strength and agility standards.

- Vertical climbs of a thirty-foot rope
- Vertical climbs of a fifty-foot ladder or wall with foot and hand holds.
- Horizontal jumps over a twenty-eight-foot span.
- Crawls beneath or slips between a ten-inch space.
- Swims in sea-state-one for one mile.

Of course, few humans can perform to any of the above limits. A more practical standard for robotic support vehicles might be the modal performance standards of the human population. But some developers may question the wisdom of applying such mediocre and languorous standards to a species with fewer constraints and greater promise.

6.2. The Control of Activity and Movement

Historically, our concept of operation of machines involves machine quiescence until a specific instruction is given by the operator to the machine to commence some pre-defined algorithm. The instruction is the initiating action that I will call here a provocation. Biological agents also respond to provocations with pre-planned algorithms, but the variety of sources of the provocation can be quite large. In addition, the relationship between any specific provocation and a particular algorithm often is acquired through experience. Prior to the acquisition of significant experience, a naive agent may respond to provocation in a generalized way that could involve simply bolting from its position. This bolting might facilitate an escape from the provocation. In this case, an observer might notice that the provocation was external to the agent, a loud sound for example. On other occasions, observers notice that agents appear to spontaneously move about with no obvious provocation. What the observers cannot notice in most of these cases, however, is the provocation from within the agent. Hunger, defined by a drop in energy reserves, is a common provocation that motivates agents. At other times, the random firing of neurons due to an accumulation of intrinsic and extrinsic sub-threshold noise is sufficient to initiate overt behavior.

7. Recapitulation

Combat operators of unmanned ground vehicles report that mobility is a serious limiting factor in their usefulness. Because of their low stature, small unmanned ground vehicles rarely can scale obstacles of heights greater than 10 inches, regularly stall on underbrush, and frequently fail to penetrate dense growths of trees, all of which admit human operators due to the human’s flexibility and multiple degrees of motion freedom. It is possible to add motion degrees of freedom to a small unmanned ground vehicle, but this creates a more difficult to solve problem of control and coordination.

Most existing unmanned vehicles are controlled by teleoperation. The human operator, usually through a joystick and radio link, directs a robot’s single degree of freedom, or multiple degrees of freedom sequentially, to execute some maneuver. Humans require intensive training, often taking years, to manage the coordination of more than one degree of freedom (for example – in playing the piano). Because of this human cognitive/performance limitation, the use of small unmanned ground vehicles with sufficient degrees of motion freedom for operation in tactical situations involving obstacle dense natural terrain will likely not be possible without competent and adaptive control processes resident on the vehicle[28]. It is to this requirement that the present effort is dedicated.

The present approach builds upon an idea that is at least several hundred millions of years old. This idea is that agent intelligence must develop from processes that promote the survival of the agent. We took this idea and first built a robot agent (Figure 1), adhering closely to existing military requirements for the Future Combat Systems (FCS) Soldier Unmanned Ground Vehicle (SUGV), but added to those requirements elements necessary (but yet insufficient) to develop an intelligent adaptive controller. The elements are multiple degrees of motion freedom, and sensors of critical events in the internal and external environments. Needed to complete the elements, and of which the present effort intends to supply, are hard-wired fixed action patterns, semi-modifiable basic reactive patterns, and the mechanisms by which our robot agent will be able to acquire mobility and survival skills. The control architecture will contain these elements and permit the acquisition of novel behavioral patterns by the robot to improve its adaptation to its environment.

Nearly all practical unmanned vehicle systems to date depend upon human decision making during mission execution. The degree of dependence is proportional to the complexities of the mission and of the operational environment. Developers have hoped to reduce human involvement by automating the required decision making processes and embedding them in the vehicles, but this makes the systems fragile under uncertainties. We intend to take the process beyond automation to permit our robotic agent to make operational decisions and learn novel behaviors using criteria related to internal state variables associated with the agent's health. We expect that this approach will retain the advantages of both independent activity and human involvement by providing the means by which the vehicle can evaluate responses to novel circumstances, and by which a human operator may become associated with certain favorable state changes of the agent, and then control the agent through biasing certain of the robot's intrinsic goals, and by aperiodic negation of the robot's selected means to those goals, rather than through an operator's constant exertion to drive the robot to the operator's objectives. This approach will result in a very different kind of an artificial agent. Because our aim with this work is to lay the essential foundation for all higher-level intelligent processes that emulate the biological, when successful we will be well-prepared to explore methods for decision making and tactical behaviors in the agent that are required for collaboration with other unmanned systems, and with humans.

[28] In addition, radio-frequency communication limitations will have negative consequences for remote control of unmanned vehicles in complex scenarios.

8. References

Blackburn, M.R. Spontaneity, Accommodation and Adaptation in Neuron Simulation. Paper presented at Neural Models of Plasticity: Theoretical and Empirical Approaches, Woods Hole, MA, April 29 - May 1, 1987.

Blackburn, M.R. and Nguyen, H.G. Learning in robot directed reaching: A comparison of methods. *Proceedings of the 1994 Image Understanding Workshop*, Monterey, CA. Nov 13-16, 1994, 781-788.

Kandel, E.R. and Hawkins, R.D. (1992) The biological basis of learning and individuality. *Scientific American*, 267, 78-86.