A Safety Integrated Architecture for an Autonomous Excavator

Conrad J. Pace, Member, IEEE, Derek W. Seward, Ian Sommerville

Abstract- This working paper presents the foundation which has been set for developing an autonomous system architecture that specifically integrates safety aspects. A hybrid architecture is proposed with particular features for safety integration. The major issue, in this paper is the clear outlining of safety and safety aspects within the operation of the system. This is done in order to aid the development of a safety case for the autonomous excavator's functionality.

I. Introduction

Any autonomous robotic system which needs to become a feasible industrial proposition, will need to comply with various safety standards. Such standards ideally require system safety to be considered from the outset of the system design. Indeed, one of the principle standards which is directly applicable to systems such as autonomous robots, IEC 61508 [1], clearly identifies a development pattern, the safety life cycle, as the basis on which systems have to be developed and validated. Yet, one of the major issues in modern autonomous systems, such as mobile robotics, is that system complexity may be high enough to give rise to substantial difficulties in assessing safety and validating the system [2]. Worse still, situations may arise where it is practically impossible to carry out the required detailed safety analysis in order to provide assurance of safety compliance [3], unless specific consideration for such an analysis is given in the design and development of the system.

Safety within industrial robotic systems has been a concern for numerous years and research in the field has been quite substantial [4]. Various standards have been published for such systems [5][6][7], these being generally targeted towards industrial robotic systems. However, the application of such standards to autonomous mobile robotic systems such as mobile excavators is not a practical proposition, for various reasons. Primarily the mobility of such prevents the isolation of the robot from humans or the wider environment. Indeed, this is not only unpractical but also impossible, since such systems l need to interact with their environment, rather than be isolated from it. As a result, such autonomous systems will need to perceive their environment, which is generally unstructured, and at best only partially known, as they carry out the action required to achieve the specified goal. This imposes a major safety concern which is generally not considered in the case of industrial robotic systems, where environments are well structured and fully known and perceptual requirements, if at all necessary, are very limited.

Furthermore, the issue of a well-structured and known environment allows for a very deterministic safety analysis to be performed in the case of industrial robots. Although still complex enough to make a safety analysis far from straight forward, such industrial robotic systems are bounded in nature by their well-defined and controlled environment. Therefore hazardous events can be, in the vast majority of cases, completely eliminated or monitored closely, and very well contained [3][8]. However, when it comes to analysing the safety implications of autonomous mobile systems in external environments, as would be the case for a robotic excavator, where there is no control of the dynamic environment, it becomes a virtually intractable problem to ensure the elimination of hazardous events. Even hazard containment becomes a very arduous task. Furthermore, the environment will only be partially observable, due to practical sensory and perceptual limitations, resulting in the added safety problem of having to deal with the uncertainty resulting from this limitation on perception.

As a result of the above constraints and the necessary need to consider safety as a major system requirement, work has been carried out to develop an architecture for an autonomous excavator. Such an architecture is intended to promote the integration of safe operational objectives with the minimum hindrance to task achievement. Here safety integration is not only concerned with ensuring that the 'internal system', including hardware, is fully functional. More importantly, it has to deal with the requirement to handle the system's limitation in observing its environment and furthermore, assess the system's ability to interact with its environment in a safe manner, i.e. it's operational safety.

A further objective of the proposed architecture is the facilitation of verification and validation procedures. This is provided for by the method in which safety is integrated into the system architecture. The consideration of

C. J. Pace is with the Department of Manufacturing Engineering, University of Malta, Malta.

D. W. Seward is with the Engineering Department, Lancaster University, U.K..

I. Somerville is with the Computing Department, Lancaster University, U.K.

verification and validation requirements as early as the initial architecture concept is essential for a product which can be assessed as safe [1][3][9].

This paper therefore is intended to illustrate such a proposed approach to safety. Section II outlines the safety implications and related problems within autonomous mobile robots, and more specifically autonomous excavators. This problem definition forms the basis on which the proposed architecture has been defined. The architecture is then explained in section III, followed by a description, in section IV, of how such a control structure should be able to deal with the safety issues identified earlier on. Finally, the conclusion to this paper outlines the progress expected through the full implementation of this architecture and how other issues in autonomous systems are expected to be tackled within the proposed control framework.

II. Safety Implications for Autonomous Mobile Robots

Safety has been defined as 'freedom from unacceptable risks/personal harm' [10]. A risk is considered to be the combined effect of the likelihood of occurrence of some undesirable event and the severity of its consequences in a given context. The issue of acceptable and unacceptable risk gives rise to various psychological and sociological issues [10], which underline the expectancies of a safe system. Such expectancies form the basis for setting safety specifications.

Considering the complex mode in which an autonomous excavator interacts with its environment, and the limitations imposed on the excavator in comprehending its surroundings justifies the need to consider safety implications. Furthermore, an autonomous excavator not only needs to avoid direct hazard sources such as obstacles, but also needs to perform its task reliably, so as not to give rise to indirect hazard sources as would be the case if it excavated a hole in the wrong place

In such circumstances managing risks becomes the major task, since full avoidance of risks is very likely to impossible. Furthermore, it is believed that the risks that do arise are greatly dependent on the uncertainty inherent in the environment and the limited ability to design a system which can react according to expectations under all situations encountered during operation. Risk management, thus becomes mainly a problem of managing the uncertainty of comprehending the environment and the means of interacting with it. Understanding such risks hence becomes a major safety goal in developing such systems.

Dhillon and Fashandi [11] have presented a clear aid to comprehending safety and risk implications within robotic systems. Yet their view is not focused on applications such as autonomous excavators where requirements for interacting in, at best, a partially structured, partially known environment, are present. The following expands the concepts to deal with mobile robots such as autonomous excavators.

A. *Hazard and Risk Analysis*

Undoubtedly, identifying the sources of hazards is a necessity in dealing with safety, and a hazard analysis is the tool for such hazard source identification. Two main groups of hazard sources may be outlined when dealing with such systems, these being internally and externally hazardous sources or events.

Internal hazard sources generally include failures occurring within the hardware and control software of the system, such as motor and sensory failure, breakdown of communication between elements of the system, battery discharge, and other such failures. These are the types of hazard sources which most safety standards, and standard safety design approaches try to deal with, by using various techniques such as fault trees analysis (FTA) and failure mode and effect analysis (FMEA). Furthermore, online diagnostic routines to cope with hardware and software failures, either through forms of redundancy involving switching over from one system component to another, or by some other form of system reconfiguration have been implemented with success [8][12][13][14].

External hazard sources, on the other hand, are of a completely different nature. Here, the hazards and hazard sources are a result of the robot's interaction with its environment, and not directly linked to any 'visible' system malfunction. The mode of interaction gives rise to complex chains of events, resulting in changes in the environment, which can themselves give rise to hazards. Such hazard sources for the case of an autonomous excavator may include various event sequences leading to collisions with various environmental entities, toppling in various situations and even more indirect hazards resulting from the modification imparted to the environment by the excavator.

Due to the nature of such hazard sources, it is not possible to eliminate the risk involved, or reduce the risk by using some statistical data of failure. A more heuristic approach needs to be applied which is mostly focused on managing the sequence of events to reduce the risks involved in the performance of the task. Furthermore, no straightforward online failure diagnostic system can be applied in this case, because there is no clearly identifiable failure, as would be the case for internal hazard management.

Although external hazard sources occur in numerous forms and modes making the task of hazard classification very arduous, it may be considered that such hazard sources occur due to either of two system deficiencies:

1. The inability to perceive the external hazard or external hazard source; this is mostly related to constraints in sensory data gathering and processing but also brings into context the notion of real-time perception.
2. The failure to react to the perceived hazard; this is considered to be mostly linked to decision making process weaknesses, i.e. the inability to generate an action which will avoid, reduce or eliminate the occurrence of a hazard source or hazard itself. Again real-time reaction plays a fundamental role in the ability of the system to deal with such external hazards and hazard sources.

Defining and analysing hazard sources must, furthermore, be put within the contest of the severity of hazard consequences Of particular importance is the requirement to determine the entities within the environment, which are at risk from hazard occurrences, particularly humans. This issue brings to the fore the necessity for considering human-robot interaction [15][16]. Aspects such as comprehending human actions and the means of communication with humans, thus become fundamental for safe operation.

B. *Safety Aspects during Development*

The need to integrate safety aspects at each development stage can be clearly seen when it comes to validating and verifying such systems. Verification looks at the compliance with the system specifications whereas validation determines how satisfactory the specifications are for the system in achieving its objective. Verification, though substantially complex for such autonomous systems, may be achieved through various techniques including formal design methods. Proving internal system integrity is largely a case of verification. However, with mobile robotic systems, where interaction with the environment plays a fundamental role in maintaining safety, validation is the main challenge

C. *Location of Safety within the System*

The requirement to consider safety at each stage of the development reinforces the need for safety to be an integral part of functional control. This includes the elements responsible for perception and action decisions. Safety harmonisation with control therefore requires not only conventional approaches such as statistical risk and reliability analysis but also, an approach for integrating safety within the 'reasoning' and 'interpreting' processes of the control system. Therefore, a control architecture which actively' ensures safety through its own actions needs to be implemented. A robotic system interacting with its environment needs to maintain safety through its behaviour, and hence through the combination of the processes of perception and action.

III. An Architectural Framework for Safety Integration

The safety requirements outlined above have led to the development of an architectural framework which encompasses all such requirements in a manner that ensures safety and dependability in the system's operation. Robot architectures tend to fall into one of two schools. Either they are hierarchical and contain an internal model or representation of the robot world, or they are reactive with behaviours that specific particular responses to particular sensory inputs. The framework presented here follows a hybrid hierarchical/reactive architecture on which both safety and control requirements are mapped. The reasons for utilising a hybrid form of architecture are delineated hereunder

- A purely reactive control architecture, although capable of reacting to the environment, can only perform rather rudimentary tasks, and more importantly, does not clearly have the capacity of analysing the consequences of its actions or of the changes in its environment [19]. Due to its very nature, such an architecture-based robot can only deal with the current situation in which it is, thus being referred to as a 'situated' system. This is considered to be unsatisfactory from a safety viewpoint, particularly if the main concern is to avoid the occurrence of hazardous situations, rather than having to deal with their occurrence.
- At the other end of the architectural spectrum, purely hierarchical control architectures do allow for the ability of consequence analysis, mainly as a result of their ability to reason at various abstract levels. Yet, such architectures may lack the same ability to react to real-time environmental changes, particularly in conditions where the environment is only partially known and unstructured. Therefore, although such systems have the ability to avoid perceived potential hazards, they do not have the potential of dealing with the eventuality of such hazard occurrences when hazard prediction fails. Furthermore, such architectures suffer from an inadequacy to deal with the uncertainties in the environment [19].

Thus, a hybrid architecture tries to derive the benefits of both purely reactive and hierarchical architectures. The features of typical hybrid architectures have been outlined by various authors [3][17][18][19] exhibiting both a reactive, behavioural component and a hierarchical, abstract reasoning component. A hybridised architecture, therefore, offers the facility of dealing with real-time aspects through its reactive layer, but is also capable of having a more deliberative form of behaviour through guidance from abstract-reasoning based, command and control layers.

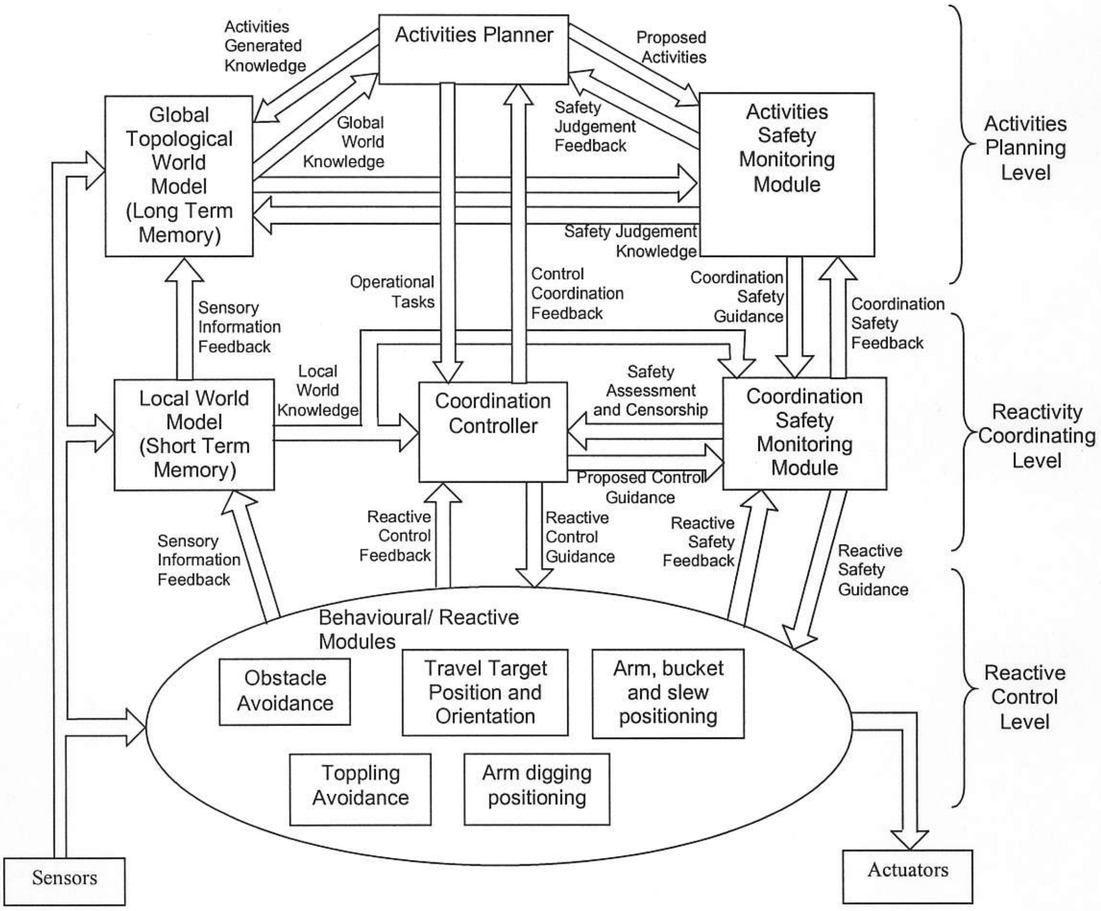
Activities Generated Knowledge
Activities Planner
Proposed Activities
Global Topological World Model (Long Term Memory)
Global World Knowledge
Safety Judgement Feedback
Activities Safety Monitoring Module
Activities Planning Level
Safety Judgement Knowledge
Operational Tasks
Control Coordination Feedback
Coordination Safety Guidance
Coordination Safety Feedback
Sensory Information Feedback
Local World Knowledge
Local World Model (Short Term Memory)
Coordination Controller
Safety Assessment and Censorship
Coordination Safety Monitoring Module
Reactivity Coordinating Level
Proposed Control Guidance
Sensory Information Feedback
Reactive Control Feedback
Reactive Control Guidance
Reactive Safety Feedback
Reactive Safety Guidance
Behavioural/ Reactive Modules
Obstacle Avoidance
Travel Target Position and Orientation
Arm, bucket and slew positioning
Toppling Avoidance
Arm digging positioning
Reactive Control Level
Sensors
Actuators

Figure 1. Layout for a Control Architecture with Safety Integration

Hence, the integration of a reactive control layer with hierarchical command features, not only provides for better task achievement when the task becomes relatively complex, but also, and most importantly in this case, it allows for the better management of safety. Such an architecture can therefore be able to avoid and manage potential hazards both by reacting to the current state of operation and also by being able to analyse consequences of its own actions.

A Three Layered Architecture

The proposed hybrid architectural framework consists of a three layered architecture, with;

i. a *lower reactive control layer*, which operates using a number of parallel running modules and which are influenced (but not directly controlled) by upper control levels,

ii. an *intermediate reactivity-coordinating layer*, which utilises a local egocentric world map for coordinating the system's actions, mainly by its influence on the lower reactive layer and,

iii. *A top activities-planning layer*, which operates at a high abstract level of reasoning, and uses a topological global world model for planning its actions.

Figure 1 illustrates the basic layout of the architecture.

This architecture has similarities to what has been proposed by Chung et al [18] but the major difference in this model is the specific mapping of safety onto the architecture. This architecture has been developed for a modified version of a JCB 801 mini excavator

The essential functionality aspects within this architecture are outlined below.

1) Low Level Reactive Control

The lowest level of control consists of a set of reactive modules all running in parallel which integrate both control and safety features. Each of these modules receives data and produces an output that is either a vector of motion, or a limitation on the possible motion. The latter may take the form of a maximum allowable operational speed. Also, certain status information is generated by specific modules and transmitted either to the upper control levels or utilised for influencing other reactive modules. The vector outputs and vector limitations are integrated into one motion output by other modules which deal with conflict resolution and

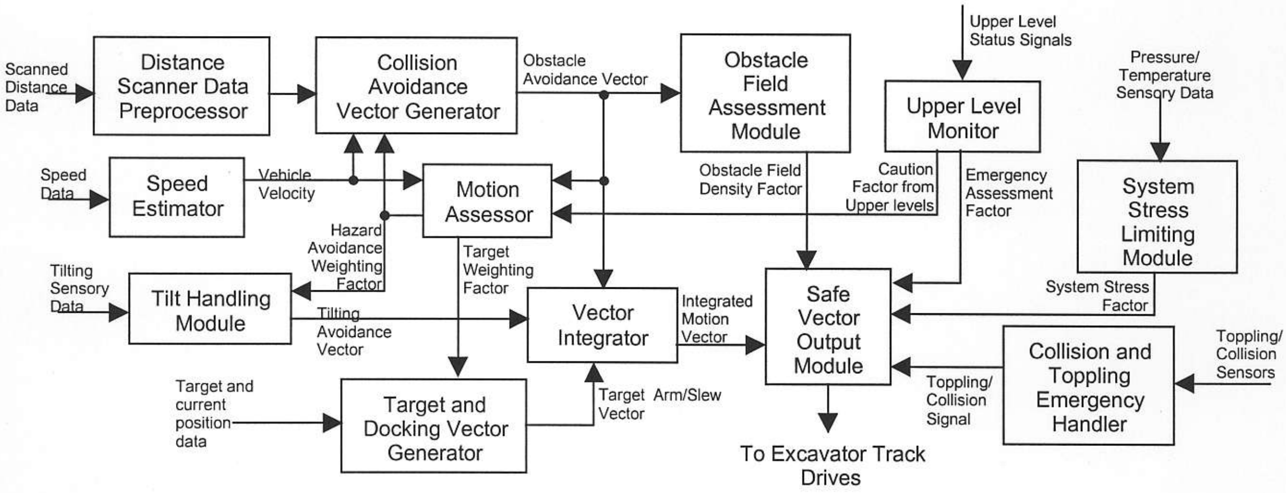
Upper Level Status Signals
Scanned Distance Data
Distance Scanner Data Preprocessor
Collision Avoidance Vector Generator
Obstacle Avoidance Vector
Obstacle Field Assessment Module
Upper Level Monitor
Pressure/ Temperature Sensory Data
Speed Data
Speed Estimator
Vehicle Velocity
Motion Assessor
Obstacle Field Density Factor
Caution Factor from Upper levels
Emergency Assessment Factor
System Stress Limiting Module
Hazard Avoidance Weighting Factor
Target Weighting Factor
Tilting Sensory Data
Tilt Handling Module
Vector Integrator
Integrated Motion Vector
Safe Vector Output Module
System Stress Factor
Tilting Avoidance Vector
Collision and Toppling Emergency Handler
Toppling/ Collision Sensors
Target and current position data
Target and Docking Vector Generator
Target Arm/Slew Vector
Toppling/ Collision Signal
To Excavator Track Drives

Figure 2. Basic Lower Reactive Control Layer Architecture for Travelling

coordination of the various motions required. This ensures that the motion vector fed to the actuator drives would be task efficient and at the same time safe.

Safety and control are closely integrated here since at a real-time decision level, safety and control requirements are indistinguishable. For example, even for task achievement, real-time avoidance of obstacles and toppling are also necessary from a purely operational viewpoint.

Figure 2 illustrates a simplified framework for the lower reactive layer for travelling. A similar layout is provided for digging. The only signals this layer receives from the upper layer with regards to travelling are target position coordinates and orientation, together with a health feedback from the upper layers and a measure indicating the level of caution to be taken by the system in its action.

For both travelling and digging, motion vectors for the excavator tracks and arm respectively, are generated for obstacle and tilt avoidance, together with target position achievement. Furthermore, a motion-assessing module is utilised for determining the balance between avoiding hazards (mainly obstacle and toppling avoidance) and achieving the required task. The motion assessor integrates different parameters such as the level of generation of obstacle and toppling avoidance vectors (which indicates the level of hazard avoidance intervention being required), together with cautionary signals resulting from the intermediate coordinate level. The balance is also affected by the current operational speed, due to the higher risks involved at higher arm and track speeds.

The integrated vector output is then fed through a safe vector output generator, which is mainly concerned with reducing the vector magnitude depending on speed limitations arising from other sources. For example, the obstacle field speed limiter, which gathers data regarding the amount of intervention required for avoiding obstacles, gives a speed limit reflecting the concentration of obstacles. Hence, if a higher obstacle avoidance intervention is detected through this module over a short period of time, reflecting a relatively high obstacle concentration, a slower operational speed will result. Speed limitation is also provided from a system stress-limiting agent, which acts in a similar fashion to Arkin's homeostatic schemas [20]. This module ensures that the system operation does not give rise to excessive pressure and temperature surges on the system actuators and power-generating unit.

Further feedback for safe vector output is obtained from the upper level-monitoring module which determines the state of the upper levels from the health signals being received. The health signals are just used as a measure to indicate the dependability of the data received from the upper levels. A similar approach is taken for the digging action.

Further to the above, various safety constraints inhibit certain actions from occurring simultaneously, such as the slewing of the excavator arm or the movement of the tracks while the bucket is in contact with the ground. The fusion of all the modules and their operations, therefore, allows for real-time safe system performance, where the effect of 'reactive' safety gradually increases and ultimately takes over from pure task achieving reactive control as hazardous operation conditions increase.

2) The Reactivity-Coordinating Level

At the next level of control within the hybrid architecture, a low level of abstract reasoning is provided. The layer operates on an egocentric view of the world. Indeed, an egocentric, local world model is used, which is rather simplistic in nature, since the interpretation and perceptual requirements at this level are relatively limited. The world model acts partly as a short-term memory of the perceived world, forming the basis for the level's operational requirements.

The basic control requirements for this level are that of dividing the high-level activities-planner tasks into more manageable subtasks, which are then to be achieved by the lower reactive layer. There is no direct control imposition by this level on the reactive layer. Instead, this level serves mostly as guidance in order to provide for more efficient use of the lower reactive layer in achieving its goals. The division of the tasks depends greatly on the perceived local world model and is directly influenced by the safety requirements of this level. Hence, if the task objective is to reach a specific location on a building site, the coordinating level may use its local knowledge to opt for a route which will directly avoid any perceived obstacles or hazardous situations. This will ensure the minimisation of potential hazardous events which may be encountered by the lower reactive layer, that would otherwise provide for longer and less efficient task achieving operations.

A major difference between this level and the reactive layer is a better ability to distinguish safety issues from control issues. This allows for safety modules to be more distinct than in the case of the lower reactive layer. The separation of safety and control functions may lead to conflict and trade-off's might be required to come up with a working solution.

The distinction between safety and control provides for the partial segregation of safety modules from control modules. As a result of this segregation, control modules are mainly concerned with the division of the task presented to them by the upper activities-planning layer. Yet they are influenced and ultimately 'censored' by the safety modules in the mode in which tasks are subdivided. This influence does not take the form of a highly abstract symbolic communication, but rather operates on a safety windowing principle, labelling perceived zones according to the level of potential risk when operating within such zones. The reason for the influential action is to allow the control modules to generate a task subdivision which is both safe and efficient. The censoring action however provides for an ultimate veto over unsafe subtasks.

Feedback from the reactive layer, indicating its mode of operation, is used both as a feedback to the coordinating level control modules to determine the level of task achievement, and by the safety modules as an indicator of safe performance. The way in which the safety modules intervene to maintain safety will therefore be influenced by the feedback from the reactive layer. Furthermore, judgements on safety made by the upper activities-planning layer are forwarded to the coordinating layer safety modules, in order to aid the safety assessment carried out at this level.

3) The Activities Planning Layer

The segregation between safety and control is even more distinguishable at this level, this being mainly due to the long-term consequence analysis provided here. Task definition for forwarding to the intermediate coordinating level is mainly done as a coordination process between control determining agents and safety determining agents. Both control and safety agents tend to influence each other, the former by indicating the required course of action for achieving the goals originally set out, and the latter by indicating what type of tasks may give rise to higher hazard risks than others. The level of communication in this case takes the form of highly abstract symbolic messages, such as a belief judgement on the safety of a specific task or a measure of the confidence with which the system can currently operate. This type of communication is very similar to the value judgements as proposed by Albus [21], and indeed the safety agents do perform as value judging agents, ensuring that a consensus is reached between task achieving objectives and risk minimisation objectives.

This layer also uses a topological mapping of the environment which takes a global rather than a local view of the world. It is not intended to act as a metric map of the excavator's environment but is primarily a knowledge base which aids in the activities planning of the system. The topological mapping consists mainly of a network of nodes and arcs, nodes representing locations whereas arcs representing paths between locations. This topological mapping strategy has been proposed for several robotic systems [18][22], although the level of abstraction of the information represented tends to vary. In this case, the knowledge representation tends to be highly symbolic in nature, as this significantly aids in the planning strategies utilised by both control and safety agents. Information stored on the nodes and arcs is represented at different strata and is operated upon as shown in figure 3.

A specific stratum represents information regarding specific goal achievement and risk concerns, such as toppling risk or collision risk when travelling along a specific path, or operating the arm during digging. Data fed into the strata is mostly determined by its safety relatedness. Aspects regarding risk assessment or uncertainty in the perceived entities at a location or path are directly fed by the safety agents to the map, through the knowledge gained from lower level feedback, from specific sensory perception routines at these higher levels, and also from pre-defined knowledge obtained prior to operation. Reading access is permitted for the control agents, but such agents are prohibited from modifying any safety related information. In this manner, the integrity of the safety information is maintained.

When it comes to determining the task plan to be carried out, both the control and safety agents will be capable of extracting information from the topological map in order to make their own assessments. In the event of travelling from node to node along an arc, the information extracted may represent a measure of obstacle concentration expected, and any past knowledge of problems in interacting with the environment. These measures will all reflect the uncertainty in the ability to arrive safely at the required location. The control agents may then take specific task decisions depending on such information. The safety agents will further integrate environmental knowledge from the topological mapping, with the current internal system state,

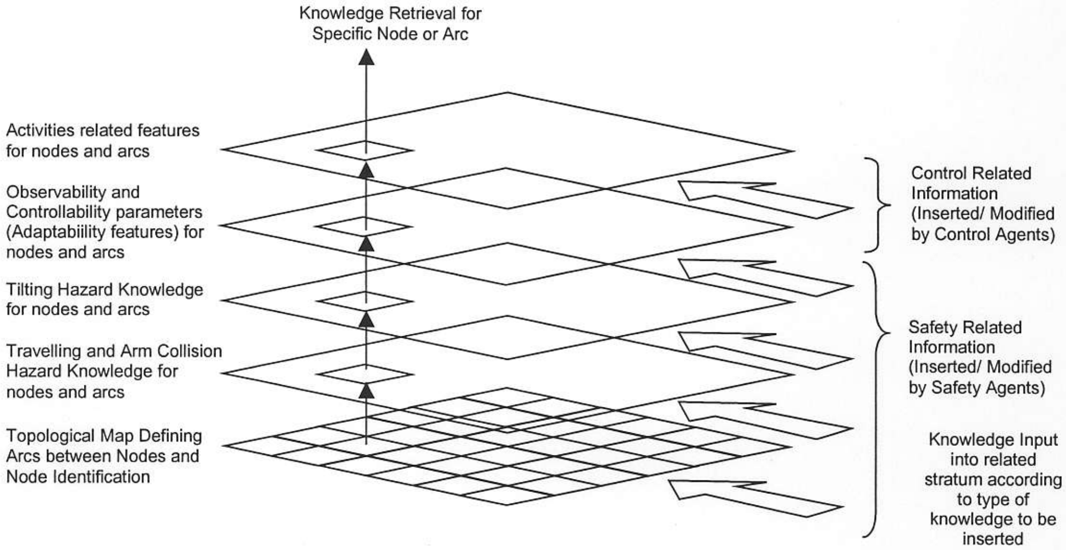
Knowledge Retrieval for Specific Node or Arc
Activities related features for nodes and arcs
Observability and Controllability parameters (Adaptabiility features) for nodes and arcs
Tilting Hazard Knowledge for nodes and arcs
Travelling and Arm Collision Hazard Knowledge for nodes and arcs
Topological Map Defining Arcs between Nodes and Node Identification
Control Related Information (Inserted/ Modified by Control Agents)
Safety Related Information (Inserted/ Modified by Safety Agents)
Knowledge Input into related stratum according to type of knowledge to be inserted

Figure 3. Structure of topological mapping strata, indicating the mode of knowledge input dependent on safety relatedeness, and knowledge retrieval dependent on node or arc.

determining a general belief value of the system's safe performance. These measures aid in determining the risk being taken by the system to perform its actions.

When the required balance between risk and goal achievement is obtained, the specified task is fed to the coordinator level. Feedback on the mode of operation of the lower levels is then used to determine task achievement from a controller's point of view, and furthermore to determine whether the risk judgement was satisfactory from the safety agents' point of view. Feedback information is also used to update the topological mapping with regards to purely task achieving features, but also, and more importantly, to map any difficulties encountered, such as tilting ground or level of obstacle avoidance required during task achievement. As a result, the topological mapping acts as a long-term memory for the system, aiding in looking at past situations and their outcomes and determining long term consequences of the actions being taken.

A safety judgement is also made on the internal system operation. This framework of assessment utilises an on line fault tree analysis approach, monitoring internal system failures. Knowledge of the state of operation of the internal system also aids in developing a proprioceptive approach as proposed in [23], where information gathered from internal sensory systems are utilised to assess the information from external sensory systems, although here, proprioception occurs at a more abstract and symbolic level.

As a result of the above actions, the upper activities planning level is not only involved in creating a plan of action to serve as a guidance for the lower levels. It will also create predictions, regarding aspects such as sensory dependability, or where the focus of attention may be required, together with a measure of the belief in the system's capacity to avoid hazardous situations, thus helping the lower levels to achieve their task more efficiently and safely.

IV. Dealing with Safety Issues within the Architecture

The proposed architecture outlined above is deemed to deal with a number of major issues in safe system design. These issues are described in the following subsections.

A. Safety Integration within different levels

As can be seen from the above, safety is built as a number of modules running in parallel with control, through the different levels. Yet, the level of integration of safety aspects with control varies according to the control level, as can be viewed in figure 4. At the lowest reactive layer, safety is substantially embedded into the reactive control itself. Safety features are almost indistinguishable from the reactive control features due to the fact that real-time control and safety decisions are indistinguishable, such as obstacle or tilt avoidance for the excavator. Furthermore, it is the combination of the safety and control modules which form individual reactive behaviours, causing the safety operational features to be intimately bound to the related control features.

On the other hand, as one moves further up the hierarchy, safety modules become more distinguishable from control, and indeed may be represented as completely separate agents at the top control layer. The logic behind this approach, as outlined earlier on, is that safety reasoning can be separated from the control counterpart. This aids in ensuring that the safety analysis performed within the system is coherent, rather than being fragmented. It allows for a more rigorous development of the safety related modules.

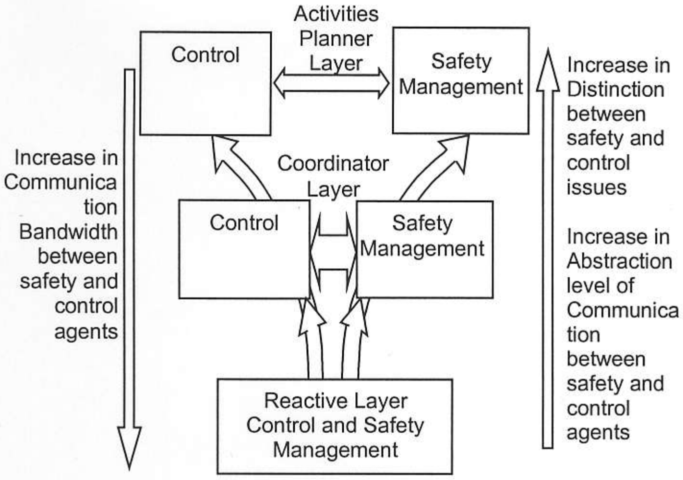
Activities
Planner
Layer
Control
Safety
Management
Increase in Distinction between safety and control issues
Increase in Abstraction level of Communica tion between safety and control agents
Increase in Communica tion Bandwidth between safety and control agents
Coordinator
Layer
Control
Safety
Management
Reactive Layer
Control and Safety
Management

Figure 4. Progressive distinction between safety and control architectural features

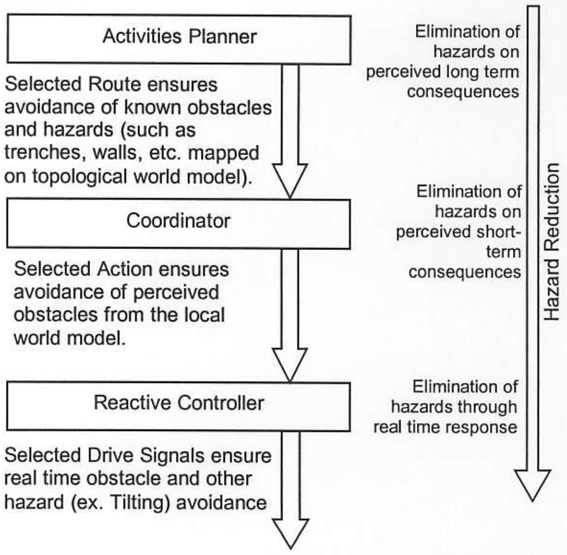
Activities Planner
Selected Route ensures avoidance of known obstacles and hazards (such as trenches, walls, etc. mapped on topological world model).
Coordinator
Selected Action ensures avoidance of perceived obstacles from the local world model.
Reactive Controller
Selected Drive Signals ensure real time obstacle and other hazard (ex. Tilting) avoidance
Elimination of hazards on perceived long term consequences
Elimination of hazards on perceived short-term consequences
Elimination of hazards through real time response
Hazard Reduction

Figure 5. Example of how hazards are handled throughout the architecture while travelling

B. Handling Safety on different Time and Spatial Scales

The division into layers is such as to provide safety on different time and spatial frames and different levels of abstraction. This is in compliance with Albus's theory of intelligence [21], although the architecture is hybrid in nature rather than of a strict hierarchical form as presented by Albus.

In the proposed architecture, the bottom most reactive level is capable of considering real time safety aspects and hazardous situations. Furthermore, the intermediate level is capable of looking at short-term consequences, avoiding potential hazards which it can predict, and thus reducing the level of safety intervention on the reactive layer. Finally, the top level can handle long term consequences of the excavator's actions and is capable of determining courses of actions which will try to minimise any hazard occurrence. This hazard reduction procedure would therefore reduce the safety workload on the coordinating level as much as is possible from a long-term consequence point of view.

Thus, the outcome of this approach is that hazard sources may be handled by each individual level to different extents, depending on the hazard mode of occurrence and the short and long-term consequences. This allows for the ability to reduce hazards further and further as task operations are passed down the hierarchical levels. For example, a potential travelling plan through a building site may be set out by the top-level activities planner. The planner will try to opt for a route which, apart from taking the excavator to the required destination in an efficient manner, also searches for a route which has been previously determined to have a low hazard occurrence probability. This hazard measurement is determined from either pre-operation acquired knowledge, or from feedback obtained via the lower levels, when passing through the same route in earlier tasks. The planner will then strike a balance between the risk being taken through the route and the efficiency of the route to arrive to destination. The planner will also, only accept a route if it is knowledgeable from information received through the lower layers, that the lower levels themselves are capable of managing the risks involved.

Having determined the route, plans and risk assessment are shifted to the coordinator level, which in turn issues target directions to the reactive layer in order to provide for the required action. The coordinator level has the function of looking at the short-term consequence of specific actions, together with being able to interpret the reactive layer's operation. Hazard avoidance interventions which are not taken by the coordinator level are then dealt with at the reactive layer level on a real-time basis. Figure 5 illustrates the mode of how such a situation is handled.

C. Inter-Layer Feedback for better management of safety

Feedback from lower levels to the levels above give an indication of the upper level's ability in interpreting the hazard potential of the actions being taken. A reactive layer which does not have to create substantial obstacle avoidance vectors, gives a clear indication to the coordinating level that the route chosen is safe enough. On the other hand, if the reactive layer has taken the excavator further away from its target position because of obstacle avoidance, this gives an indication of the upper level's inability to select a safe route. Similarly interventions recorded by the coordinating level, give an indication of the extent at which the top planner level has been able to interpret safety and risk aspects. These feedback recordings aid both the coordinating level and activities planning level to reassess their safety management abilities, potentially requiring the levels to put more emphasis on safety and hazard avoidance rather than on task achievement. In this manner, task achievability is still possible, and safety is not hindered.

D. *Environmental Interaction Uncertainty*

The guidance provided by each upper level to the level below it and the feedback provided by each lower level to the level above it, also helps in managing the uncertainty present in the system's ability to observe its environment and to control its interaction in a safe manner. Feedback obtained from the lower levels can help in the perceptual abilities of the upper levels. Perceiving a path of action to be safe at the upper activities planner level, only to receive feedback indicating substantial intervention by the lower levels in order to avoid hazards, clearly indicates that the perception of a safe path of action was a misjudgement and correction of the interpretation is required. This would result in a higher perceived uncertainty level, leading to a more tentative operational approach. Conversely, if feedback received from the intervention of the lower layers, complies with the expectations set by the upper layer, then this will reduce the perceived uncertainty in the activities planner level to observe and interpret its environment. Consequently, an indication of high confidence in the system operation would result from the modules at each level having to intervene less and less to maintain safety

The management of uncertainty directly reflects the system's ability to manage the risks involved in interacting with the excavator's surroundings. The feedback obtained reflects the action and perception abilities of the system, and counter measures can be taken to ensure that the level of interaction between system and environment is within the underlying systems' information limitations.

E. *Diversity and Redundancy*

As mentioned above, the hybrid architecture provides for the possibility of dealing with safety problems at different levels starting from a highly abstract, long term consequence assessment at the upper most level, to a real time, reactive safety maintenance at the bottom most level. The gradual reduction and elimination of potential hazards as tasks move down the hierarchy, provides for an inherent safety redundancy, which is based on completely diverse control principles and which can handle safety problems in very different manners. The diversity in the control principles is mainly due to the diverse operational requirements for each level and the different levels of abstraction at which each level operates. Consequently, hazards and hazardous situations not detectable at a specific level, may still be managed by the other levels, maintaining overall safety integrity.

Furthermore, the fact that each level is relatively independent from the other levels, allows it to continue operating even in the event of an other level failing, thus still securing safety, albeit within certain constraints. If for example, the top activities planner fails, the intermediate level may still be able to ensure that the system will operate safely, although the overall system will not be able to cater for long term consequences of the actions taken. Similarly, if the intermediate level fails, the lower reactive level will still be able to maintain the bare safety requirements for real-time safety maintenance. Yet, it will obviously, not be able to measure action consequences, and will definitely inhibit the possibility of the excavator achieving its operational task. Still, a graceful system failure and fail safe situation can be achieved, in a worst case scenario, due to this independence between levels.

F. *Development and Validation Aspects*

The mode of integrating safety requirements within the different layers as presented here, also allows for a better framework for system development within the context of the proposed safety life cycle [1]. The architectural division itself serves as a basis of managing the development of the components in order to aid the validation and verification process of development. The relative independence of the layers provides for the ability to develop, validate and verify layers individually. The inherent temporal and spatial divisions provide a basis for development structuring. Furthermore, a safety case for the excavator may be produced, based on parallel safety arguments generated from the individual layers in dealing with the same hazards in different modes. In addition, integration testing requirements may be less demanding than if the levels where significantly dependent on each other.

Further to the above, the form of diversity and redundancy inherent in the architecture allows for safety integrity to be distributed between all the levels, allowing for a higher overall system safety integrity from lower safety integrity architectural components. The diversity and redundancy aspect also greatly enhances the ability to utilise AI technology for higher abstract reasoning safety agents. Even though AI agents may only be considered to be of a low safety integrity [1], general safety is still provided for, by the fusion of the safety outcome of each level. Furthermore, the segregation between safety and control, as one moves up the architecture, allows for different safety integrity levels to be allocated to control and safety modules, allowing for different levels of development rigour as required by the allocated integrity level.

V. CONCLUSION

This paper proposes a method by which environmental interaction can be managed within system safety limits. In systems such as an autonomous excavator operating in its complex environment, exhaustive definition of the environment at the requirements definition stage of development becomes a daunting task, if not unrealisable. The architectural division into layers and the underlying division of safety, aids in defining what environmental features will be required to be perceived within the different levels, greatly facilitating the environment definition at the specification stage of development. Furthermore, the issue of dealing with the uncertainty of interpreting the environment ensures that the excavator is aware of its limitations in observability and controllability and utilises feedback from what it perceives to determine how dependable its interaction with the environment is.

The layered division and the level independence should also provide for better hazard management through the system's ability to visualise hazards on different spatial and temporal scales. This mode of defining safety accommodates a representation of safety 'awareness' through its upper levels, and a safety 'instinct' through its reactive level. The layered division also aids in avoiding over-reliance on a specific level or a specific module for ensuring safe operation, allowing for a more tractable design, particularly when considering issues of verification and validation.

The architecture proposed here also tries to emphasise the issue of transparency in the mode of control, i.e. the ability to trace out how a specific action is taken and how the safety modules are capable of handling the related safety problems. Transparency or traceability, is a major requirement for validating and verifying safety-related systems. Systems that are required to operate safely and within specific safety integrity levels, should not exhibit unexpected reactions to certain sensory stimuli, and indeed, strict representations such as formal methods may be required [1]. Yet, traceability and a clear comprehension of how the system determines a specific action is what is lacking in purely reactive systems such as the subsumption architecture [24] where behaviours are said to be emergent in nature. Although reactivity does give rise to robustness, an approach is required which allows traceability. Thus, the hybrid nature and distribution of the architecture should ensure a clearer representation of the interactions within the system, and further provides for the required robustness for safety.

A. *Current and Future Development Stages*

Currently, the architecture is in the process of being developed for the converted mini-excavator, initially on a simulation platform, with the scope of identifying and delineating the specific safety concerns for each level. The major safety and control features within each level have already been identified, and detailed safety and control models has been created for certain parts of the architecture. Further detailed analysis into the required interaction between controller and safety modules still needs to be, and is being, carried out. In this respect, the issue of safety 'awareness' can be further developed to represent the means by which the top activities-planner safety agents interpret their role. It is envisaged that the safety case [1][3][9], which develops the arguments used to validate system safety, may be directly integrated within the reasoning pattern of the top level safety agents. In this manner they can generate a more global view of the system's safety integrity by determining the validity of the safety case arguments themselves.

Further to the above, the project's aim is to focus mainly on the functional failure issues, outlining the excavator's ability to perceive its environment and act on it in a safe manner. Other 'internal' failure issues, particularly random hardware failures, are not considered to be a major concern within the scope of this work. Yet, they are still dealt with to a certain extent through the use of an internal failure-monitoring element present within the safety agents at the activities planner level. The inclusion provides for better handling of system reliability.

The handling of uncertainty as a form of managing risk resulting from the lack of knowledge in environmental interaction, will be directly embedded in the mode in which the individual modules and levels handle information. This is expected to be achieved through the utilisation of belief measures, mapped onto perceptual information and internal system state behaviour. The belief measure representation will inherently ensure that processing of information will cater for the underlying information limitations and accuracy.

Although there has been mention of the safety implications of having such systems interact with humans, consideration of such issues has not yet been given due attention. Indeed, such aspects give rise to various perceptual and communication requirements, some occurring at a more symbolic level than others. Yet, it is felt that the architecture should be able to accommodate such human interaction requirements, by being able to spread both the resulting perception and action effects across the levels, depending on the reactivity or reasoning level required. Work, in this area, though, needs to be carried out to verify such implications.

Learning is another issue which needs to catered for in the long term. Indeed, various arguments may be presented in favour or against the inclusion of learning within a safety architecture. The major advantage of learning is that the system becomes capable of coping with situations which were not originally defined. The ability to learn may therefore be considered as the ability to increase system robustness and as a result, allow for a better safety management [25]. Yet, there is no clear means of guaranteeing that the autonomously learnt perception–action coupling results in a safer behaviour, unless the learning approach is not properly monitored by some safety agent. This then presents further problems regarding the level of knowledge available to the safety-assessing agent in the first place.

REFERENCES

[1] International Electrotechnical Commission 'IEC 61508 – Functional Safety: Safety-Related Systems', Parts 1 to 7.

[2] Gaskill S.P., Went S.R.G., 'Safety Issues in Modern applications of Robots' , Reliability Engineering and Systems Safety, Vol 53 No. 3, Sep 1996, pp301-307

[3] National Advanced Robotics Research Centre, 'Safety and Standards for Advanced Robots – A First Exposition', Report ARRL.92.009, July 1992

[4] Graham, J. H., editor 'Safety, reliability, and human factors in robotic systems' New York : Van Nostrand Reinhold, 1991.

[5] Health and Safety Executive 'Industrial Robot Safety', Report HS/G 43, 1995

[6] ANSI/ Robotics Industries Association, 'American National Standards for Industrial Robots and Robot Systems – Safety Requirements', R15.06-1986, Ann Arbor, MI, 1986

[7] BSR/ Robotics Industries Association, 'Proposed American National Standard for Industrial Robots and Robot Systems – Guidelines for Reliability Acceptable Testing', Ann Arbor, MI, 1993.

[8] Visinsky M.L., Cavallaro J.R., Walker I.D., 'Robotic fault detection and fault tolerance: A survey' , Reliability Engineering and System Safety, Vol 46, No. 2, 1994, pp139-158.

[9] Storey Neil, 'Safety Critical Computer Systems', Addison Wesley

[10] David Blockley, editor, 'Engineering Safety', McGraw-Hill, 1992.

[11] Dhillon B.S., Fashandi A.R.M., 'Safety and Reliability Assessment Techniques in Robotics', Robotica, Vol 15, 1997, pp701-708

[12] Visinsky M.L., Cavallaro J.R., Walker I.D., 'Robotic Fault Tolerance: Algorithms and Architectures', in 'Robotics and Remote Systems for Hazardous Environments',Jamshidi M. and Eicher P.J. Eds., Prentice Hall, 1993, pp 53 – 73.

[13] Visinsky M.L., Cavallaro J.R., Walker I.D., 'A Dynamic Fault Tolerance Framework for Remote Robots', in IEEE Transactions on Robotics and Automation, Vol 11m No. 4, August 1995, pp 477- 490

[14] Drotning W., Wapman W., Fahrenholtz J., Kimberly H., and Kuhlmann J., 'System Design for Safe Robotic Handling of Nuclear Materials', in Proceedings of the 1996 2nd Speciality Conference on Robotics for Challenging Environments, June 1996 pp 241-247.

[15] Wilkes D. M., Alford A., Pack R. T., Rogers, T., Peters, R. A. II; Kawamura, K. 'Toward Socially Intelligent Service Robots', Applied Artificial Intelligence, v 12, n 7-8, Oct-Dec 1998, p 729-766.

[16] Wilkes D. M., Alford A., Cambron M. E., Rogers T. E., Peters R. A., Kawamura K, 'Designing for human-robot symbiosis', Industrial Robot, v 26, n 1, 1999, p 49-58.

[17] Kawamura K., Pack R. T., Bishay M., Iskarous M., 'Design Philosophy for Service Robots', Robotics and Autonomous Systems, v 18, 1-2, Jul 1996, p 109-116.

[18] Chung J., Ryu B.S., Yang H .S. 'Integrated Control Architecture based on behaviour and plan for mobile robot navigation', Robotica, Vol 16, 1998, pp. 387-399.

[19] Arkin R.C., 'Behaviour-Based Robotics', MIT Press, 1998.

[20] Arkin R.C., 'Homeostatic Control for a Mobile Robot: Dynamic Replanning in Hazardous Environments', Journal of Robotic Systems, Vol. 9, No. 2, March 1992, pp. 197-214.

[21] Albus J.S., 'Outline of a Theory of Intelligence', IEEE Transactions on Systems, Man and Cybernetics, Vol. 21, No. 3, May/June 1991, pp.473-509.

[22] Nehmsow U., 'Self-Organisation and Self-Learning Robot Control', IEE Colloquium (Digest), 026, 1996

[23] Solomon R., 'Improving the DAC Architecture by using Proprioceptive Sensors', in R. Pfeifer, B. Blumberg, J.A. Meyer and S.W. Wilson (Eds), 'From Animals to Animats 5: Proceedings of the 5th International Conference on Simulation of Adaptive Behavior, pp 331-339, MIT Press, Cambridge, MA.

[24] Brooks R. A, 'Intelligence Without Representation', Artificial Intelligence 47 (1991), 139-159.

[25] Donnart J.Y., Meyer J.A., 'A Hierarchical Classifier System Implementing a Motivationally Autonomous Animat', Proceedings of the 3rd International Conference on Simulation of Adaptive Behaviour, Brighton, UK, August 1994, pp 144-153.